



L'emphase dans le discours spontané anglais : corrélats acoustiques et prosodiques

Sophie Herment-Dujardin

► To cite this version:

Sophie Herment-Dujardin. L'emphase dans le discours spontané anglais : corrélats acoustiques et prosodiques. Linguistique. Université de Provence - Aix-Marseille I, 2001. Français. NNT : . tel-00010114

HAL Id: tel-00010114

<https://theses.hal.science/tel-00010114>

Submitted on 12 Sep 2005

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

UNIVERSITE AIX-MARSEILLE I – Université de Provence
U.F.R. LAG-LEA – Département d’Etudes du Monde Anglophone
Formation doctorale LANGAGE et PAROLE

THESE
pour obtenir le grade de
DOCTEUR DE L’UNIVERSITE AIX-MARSEILLE I

présentée et soutenue publiquement
le 15 décembre 2001
par

Sophie Herment-Dujardin

***L’emphase dans le discours spontané anglais :
corrélats acoustiques et prosodiques***

en deux volumes accompagnés d’un CD ROM : VOLUME 1

sous la direction de
Monsieur Daniel HIRST
Directeur de recherche, CNRS, UMR 6057
Institut de phonétique, Laboratoire Parole et Langage - Université de Provence

JURY

Monsieur Jean-Louis Duchet, Professeur à l’université de Poitiers
Monsieur Daniel Hirst, Directeur de Recherche CNRS à l’université de Provence
Monsieur Renaud Méry, Professeur à l’université de Provence
Monsieur Peter Roach, Professeur à l’université de Reading, Angleterre
Monsieur Michel Viel, Professeur à l’université de Paris IV la Sorbonne

A mon fils,
A mes parents.

REMERCIEMENTS

De nombreuses personnes m'ont aidée tout au long de ce travail. Je remercie chaleureusement

Daniel Hirst,

qui a accepté de me suivre et de me diriger. Je lui suis reconnaissante pour sa disponibilité et sa gentillesse, pour les nombreux conseils éclairés et tout le savoir qu'il transmet si modestement.

Michel Viel, Jean-Louis Duchet, Peter Roach, et Renaud Méry,

qui m'ont fait l'honneur d'accepter de faire partie de mon jury.

Michel Ginésy,

qui m'a fait découvrir et aimer la phonétique. Ses qualités intellectuelles, couplées à d'immenses qualités humaines, forcent mon admiration. C'est un privilège et un honneur mais surtout un grand bonheur que de travailler avec lui. Merci pour tous les encouragements que j'ai reçus, pour toutes les préparations de cours et d'examens, pour les corrections. Merci de m'avoir appris et de m'apprendre encore tant de choses.

André Lipcey et Philippe Dominique,

qui m'ont également beaucoup appris quand j'étais leur étudiante puis leur collègue, et m'ont aussi transmis leur amour de la phonétique.

Vicki Briault,

pour ses encouragements et pour les enregistrements. Tu as su me remotiver au bon moment. Je t'en suis reconnaissante.

Les collègues anglophones du département d'anglais de l'université d'Aix-en-Provence,

qui ont accepté de se prêter aux enregistrements ou à l'écoute des enregistrements, ainsi qu'à tous les auditeurs.

Nicolas Giaramidaro et Georges Bourrat, pour leur aide technique, leur disponibilité et surtout leur gentillesse.

L'équipe des anglicistes du labo de phonétique (The English Prosody Group), et plus particulièrement *Caro, Cyril et Nadine,* grâce à qui la croisière de la thèse fut ponctuée de bonnes rigolades, mais aussi *toute l'équipe du laboratoire Parole et Langage* et tout spécialement

Isabelle Marlien, pour ses relectures et ses précieux conseils éditoriaux, sa gentillesse, et ses encouragements, et

David Azulay, qui a fait preuve d'une grande disponibilité pour m'aider à générer les arbres de classification.

Paul Boersma,

qui m'a permis d'utiliser le logiciel PRAAT. Je souhaite lui rendre hommage et lui exprimer mon admiration.

J'ai aussi eu la chance de bénéficier d'un immense soutien de ma famille et de mes amis, sans lequel ce travail ne serait pas.

Je tiens à adresser mes plus sincères remerciements à

Bertrand, mon mari,

merci de m'avoir supportée (dans tous les sens du terme). Tu es mon plus grand fan et tu as toujours su trouver les mots ou les gestes salvateurs. Merci d'être le père le plus merveilleux du monde pour notre fils. Merci pour tout ton amour. Je n'aurais jamais réussi sans toi.

Gautier, mon fils,

mon p'tit bonhomme. Ta gaité, tes éclats de rire et tes yeux filous m'ont été très précieux. A la réflexion, je crois que tu avais tout compris depuis bien longtemps. Je te dois beaucoup.

Mes parents,

ma mère m'a transmis son amour des langues, des lettres, des livres et des mots ; mon père m'a fait aimer le côté scientifique et rigoureux des choses ; tous deux m'ont appris la persévérance et la patience. Ce travail est le fruit de ce savant mélange.

Ma gratitude va bien au-delà de cela ; tous deux m'ont aidée à leur façon tout au long de ce travail. Je remercie mon père pour les précieux conseils techniques et pour les relectures sur la partie statistique. Je remercie ma mère d'avoir subi mes humeurs sans un mot et d'avoir été toujours présente. Votre soutien sur tous les plans est inestimable et vous êtes des grands-parents formidables.

Mes beaux-parents,

qui me soutiennent et m'encouragent depuis fort longtemps déjà et qui sont aussi des grands-parents formidables.

Hervé, mon frère,

dont j'ai toujours été admirative et dont l'exemple a toujours été une motivation,

Christine,

pour son soutien sans faille et sa disponibilité, et mes neveux Agathe et Thomas.

Merci à vous quatre pour tous vos encouragements. Les moments de bonheur tous ensemble furent mes bouffées d'oxygène.

Mes grands-mères, avec tout mon amour.

Cath, Od, Gilles et Mélodie, et tous mes amis, en particulier

Claire, avec qui j'ai pu partager les frustrations, les angoisses et les satisfactions qui hantent la vie d'une thésarde ;

Laetitia, pour ses encouragements, ses relectures et ses commentaires avisés ;

Fly, pour son aide et ses conseils d'informaticien dévoué.

Enfin, j'ai une pensée toute particulière pour *mes grands-pères*.

Table des matières

VOLUME 1

TABLE DES MATIERES	3
TABLE DES ILLUSTRATIONS	8
INTRODUCTION	11
CHAPITRE 1 - L'EMPHASE : DEFINITIONS ET ETUDES ANTERIEURES	16
<i>1.1 Définitions.....</i>	<i>18</i>
1.1.1 Définitions générales.....	18
1.1.2 Focus, topique, et information nouvelle.....	22
1.1.2.1 Focus, présupposition et information.....	25
1.1.2.2 Focus large et focus étroit	28
1.1.3 Contraste.....	31
1.1.4 « accent », « stress » et « prominence »	34
1.1.4.1 « Stress-as-loudness ».....	35
1.1.4.2 « Stress-as-accent »	37
1.1.4.3 « Stress-as-rhythm »	38
1.1.4.4 « Stress and non-stress accent »	39
1.1.5 Conclusion.....	39
<i>1.2 Etudes expérimentales.....</i>	<i>41</i>
1.2.1 Emotion.....	41
1.2.1.1 Paramètres acoustiques et attitudes.....	41
1.2.1.2 Contours intonatifs et émotions ou attitudes	43
1.2.1.3 Travaux récents.....	46
1.2.2 « Stress » et proéminence	48
1.2.2.1 Durée, intensité et fréquence fondamentale	48
1.2.2.2 Déclinaison de la F0 et accent prénucléaire	50
1.2.2.3 Perception et tilt spectral	52
1.2.3 Emphase.....	54
1.2.3.1 Coleman	54
1.2.3.2 Paramètres acoustiques.....	56
1.2.3.3 Variabilité entre les locuteurs.....	64
<i>1.3 Conclusion</i>	<i>67</i>

CHAPITRE 2 - BASE DE DONNEES ET METHODE.....69

2.1	<i>La base de données.....</i>	71
2.1.1	Bases de données existantes	72
2.1.2	Choix de la variété d'anglais	74
2.1.3	Constitution de la base de données.....	74
2.2	<i>Méthode.....</i>	79
2.2.1	L'emphase : phénomène catégoriel ou continu ?.....	79
2.2.2	Instructions et présentation	83
2.2.3	Tests perceptuels	85
2.2.3.1	Méthode	85
2.2.3.2	Commentaires	87
2.2.3.3	Résultats des sondages et test de perception	90
2.3	<i>Traitement du signal acoustique.....</i>	94
2.3.1	Numérisation	94
2.3.2	Etiquetage du corpus	94
2.3.2.1	PRAAT.....	95
2.3.2.2	Passages et segments	96
2.3.2.3	Mots.....	98
2.3.2.4	« Prominence »	104
2.3.3	Transcription phonétique	106
2.3.3.1	Mise en forme des fichiers	106
2.3.3.2	Corrections manuelles	110
2.4	<i>Résumé et conclusion</i>	113

CHAPITRE 3 - PARAMETRES ACOUSTIQUES ET PROSODIQUES..... 115

3.1	<i>Fréquence fondamentale.....</i>	118
3.1.1	Comment analyser la fréquence fondamentale ?	118
3.1.1.1	F0 brute	118
3.1.1.2	MOMEL (MOdelisation MELodique) et PRAAT	123
3.1.1.3	Corrections manuelles	125
3.1.2	Pré-analyses et données pertinentes	131
3.1.2.1	F0 maximale et minimale	132
3.1.2.2	Syllabe, mot ou pied ? Une alternative : l'unité tonale.....	132
3.1.2.3	Temps.....	138
3.1.2.4	Contour intonatif général : pente, moyenne et écart-type.....	140
3.1.3	Résumé partiel.....	141
3.1.4	Découpage en unités tonales et intonatives.....	142
3.1.4.1	Terminologie.....	143
3.1.4.2	Frontières d'unités intonatives	143
3.1.5	Résumé	145

3.2	<i>Durées</i>	148
3.2.1	Méthode	148
3.2.1.1	Niveau segmental ou supra segmental ?	148
3.2.1.2	Niveau supra segmental : Campbell	149
3.2.2	Paramètres	150
3.3	<i>Intensité</i>	152
3.3.1	Intensité globale, fiabilité et parole naturelle	152
3.3.2	Pré-analyses	153
3.3.3	Intensité globale et paramètres	156
3.3.4	Résumé	157
3.4	<i>Pauses</i>	159
3.4.1	Différents types de pauses	159
3.4.1.1	Pauses et unités intonatives	161
3.4.1.2	Pauses et hésitation	162
3.4.1.3	Pauses et emphase	163
3.4.1.4	Ambiguïtés et prise de souffle	164
3.4.2	Nombre de pauses	165
3.4.3	Durée des pauses	171
3.4.4	Pauses emphatiques : avant ou après ?	173
3.4.5	Résumé et conclusion	175
3.5	<i>Cas particuliers</i>	177
3.5.1	Baisse de l'intensité	177
3.5.2	Attaques des mots et coups de glotte	178
3.5.3	Qualité de la voix	182
3.6	<i>Résumé et conclusion</i>	183
 CHAPITRE 4 - ANALYSE STATISTIQUE		184
4.1	<i>CRUISE : généralités</i>	187
4.1.1	Fichiers d'entrée	187
4.1.2	Fonctionnement du programme	188
4.1.3	Fichiers de sortie	188
4.2	<i>Paramètres</i>	190
4.2.1	Variable dépendante (d)	190
4.2.2	Durée	191
4.2.3	Fréquence fondamentale	192
4.2.4	Intensité	195
4.2.5	Pauses	196
4.2.6	Amplitude totale	197
4.2.7	Validation	198
4.2.8	Locuteurs et passages	199

4.2.9	Résumé : fichier descriptif.....	199
4.3	<i>Essais préliminaires</i>	202
4.3.1	SE rule	202
4.3.2	Valeurs par défaut	203
4.3.3	Coûts de classification erronée	204
4.3.4	Nouveaux coefficients d'emphasis.....	210
4.3.5	Résumé et conclusion : discussion sur l'aspect catégoriel de l'emphasis	213
4.4	<i>Analyses</i>	216
4.4.1	Remarques préliminaires	216
4.4.2	Corpus primaire.....	220
4.4.3	Types de discours.....	227
4.4.4	Locuteurs	230
4.5	<i>Résumé et conclusion</i>	234
 CHAPITRE 5 - ETUDE PERCEPTIVE SUR SIGNAUX MANIPULES		235
5.1	<i>Test de perception</i>	237
5.1.1	La resynthèse vocale : principes généraux.....	237
5.1.2	Phrases-sources	266
5.1.3	MBROLA	242
5.1.4	Manipulation des paramètres acoustiques et prosodiques.....	246
5.1.4.1	Durées segmentales	246
5.1.4.2	Fréquence fondamentale.....	247
5.1.4.3	Pauses.....	247
5.1.5	Délexicalisation.....	248
5.1.6	Mise en place du test.....	249
5.2	<i>Résultats</i>	251
5.2.1	Phrases originales et phrases synthétiques.....	251
5.2.2	Pauses	254
5.2.2.1	Omission d'une pause.....	254
5.2.2.2	Ajout d'une pause.....	256
5.2.2.3	Conclusion	256
5.2.3	Durées	256
5.2.3.1	Niveau du mot.....	257
5.2.3.2	Mots adjacents	258
5.2.3.3	Phonèmes	260
5.2.3.4	Conclusion	261
5.2.4	Fréquence fondamentale	261
5.2.5	Sémantisme.....	265
5.2.5.1	Mots au sémantisme fort	265
5.2.5.2	Mots « neutres ».....	267
5.3	<i>Résumé et conclusion</i>	271

5.3.1	Résumé	271
5.3.2	Conclusion	272
CONCLUSION		275
RÉFÉRENCES BIBLIOGRAPHIQUES		283
INDEX.....		300
<i>Index des notions et mots-clés.....</i>		<i>301</i>
<i>Index des auteurs.....</i>		<i>304</i>

VOLUME 2 : ANNEXES ET CD ROM

TABLE DES ANNEXES	308
ANNEXES	309
CONTENU DU CD ROM	391

Table des illustrations

Figures

Figure 1 : Question Time	97
Figure 2 : P3.1S06	99
Figure 3 : P3.1S08	99
Figure 4 : P2.6S08	100
Figure 5 : P2.1S04	100
Figure 6 : P1.1S01	101
Figure 7 : P1.5S04	101
Figure 8 : P1.5S02 : fichier son et fichier TextGrid.....	103
Figure 9 : P1.5S02	105
Figure 10 : fichier .mtx P1.1S04.....	108
Figure 11 : fichier .mtx 'enrichi' (mtx+phon) P1.1S04	109
Figure 12 : P1.7S04, F0 brute	119
Figure 13 : P1.1S08, F0 brute	120
Figure 14 : P1.2S03, F0 brute	121
Figure 15 : P2.4S03, F0 brute	121
Figure 16 : P1.9S05, F0 brute	122
Figure 17 : P2.3S01, F0 brute	122
Figure 18 : P3.3S05, courbe brute en haut, courbe lissée en dessous	124
Figure 19 : P2.4S02, courbe brute en haut, courbe lissée en dessous	125
Figure 20 : courbes de F0 brute avec plafond à 600 Hz des segments P1.4S04, P1.4S05, P1.4S06 et P1.4S08, tous prononcés par PM	127
Figure 21 : courbes de F0 brutes avec plafond à 350 Hz pour les segments P1.4S04 et P1.4S06 prononcés par PM.....	128
Figure 22 : courbes de F0 brute pour P2.3S01 : la courbe du haut a un plancher à 75 Hz, la courbe du bas à 140Hz	129
Figure 23 : segment P1.3S01	130
Figure 24 : P1.8S04 : ton creusé sur « principe »	135
Figure 25 : P3.1S09	136
Figure 26 : P2.6S08 : TextGrid avec trois « tiers » : words, UT, UI.....	146
Figure 27 : P1.3S05 et P1.3S08 : son, TextGrid et « intensity tier ».....	155

Figure 28 : P2.5S03, son , TextGrid et « intensity tier »	178
Figure 29 : P1.3S03 : longueur du groupe consonantique initial de « scooters ».....	179
Figure 30 : P1.2S05 : tenue du phonème initial de « different »	180
Figure 31 : P1.8S08 : attaque de « actually »	181
Figure 32 : P3.1S03 : coup de glotte devant « actual ».....	182
Figure 33 : amplitude totale	197
Figure34 :emph-2-0, arbre correspondant au fichier emph-2-0.out	217
Figure 35 : arbre de régression pour tout le corpus primaire.....	222
Figure 36 : arbre de régression pour le fichier emph-0-8, avec 1SE.....	226
Figure 37 : P1.1S01, fichier son + TextGrid avec phonèmes en alphabet SAMPA.....	243
Figure 38 : P1.3S01.pho.....	245
Figure 39 : son, TextGrid et « intensity tier » pour P1.3S01 : « on Christmas Day »	253

Graphiques

Graphique 1 : corrélation entre le pourcentage d'emphase et le pourcentage de degré d'emphase pour le sondage 1	91
Graphique 2 : corrélations entre le pourcentage d'emphase et le pourcentage de degré d'emphase pour les sondages 1, 2 et 3 confondus	92
Graphique 3 : nombre de pauses par locuteur	166
Graphique 4 : distribution des pauses avec des pourcentages en abscisses	167
Graphique 5 : pourcentage de pauses marquant des frontières d'unités intonatives par locuteur.....	168
Graphique 6 : pourcentage de pauses marquant des hésitations par locuteur	169
Graphique 7: pourcentage de pauses emphatiques par locuteur	169
Graphique 8 : durées moyennes en millisecondes des trois types de pauses	172
Graphique 9 : durées moyennes en millisecondes des trois types de pauses par locuteur	173
Graphique 10 : parmi les pauses emphatiques, pourcentage de celles qui se trouvent avant et de celles qui se trouvent après un mot emphatique, par locuteur.....	175
Graphique 11 : pourcentage cumulé d'emphase pour toutes les séries de stimulus	273

Tableaux

Tableau 1 : nombre de pauses emphatiques, de pauses précédant un mot emphatique, de pauses suivant un mot emphatique	174
Tableau 2 : fichier descriptif emph.dsc.....	201
Tableau 3 : ordre aléatoire de présentation des stimulus du test perceptif sur signaux manipulés pour les groupes 1 et 2.....	250

Introduction

L'enfant dans le ventre de sa mère entend ce qui se dit autour de lui mais l'enveloppe dans laquelle il se trouve joue le rôle d'un filtre faisant que l'enfant ne perçoit pas les mots, mais la prosodie. C'est grâce à la prosodie qu'il reconnaît sa langue maternelle et c'est pour cela que, lorsqu'il va apprendre une autre langue, c'est sa prosodie qu'il aura le plus de mal à modifier, celle-ci étant ancrée chez lui pratiquement depuis sa conception. Ce phénomène nous semble fascinant, et c'est une des raisons pour lesquelles nous avons souhaité étudier la parole et plus précisément la mélodie de la langue. Tout petit, lorsque l'enfant ne sait pas encore parler et commence à peine à comprendre les mots, il sait si on le gronde ou non, il est rassuré si on lui parle doucement. Il semble capable de percevoir du sens sans comprendre les mots. Il sait dès le plus jeune âge reconnaître une mise en relief, quelle qu'elle soit, dans la bouche de la personne qui lui parle. Cela signifie qu'au-delà des mots, la prosodie joue un rôle crucial dans la parole.

Nous venons de parler de mise en relief et nous nous sommes intéressée plus précisément à la notion d'emphase car il semble qu'elle fasse partie de ces acquis que personne ne remet en question. Certes, l'emphase est un phénomène difficilement définissable, pourtant il semble *a priori* facile de pointer une emphase dans le discours de quelqu'un. C'est pourquoi nous avons souhaité faire une étude linguistique de l'emphase en nous concentrant plus particulièrement sur les aspects acoustiques et prosodiques.

Comme nous venons de le mentionner, il est très difficile de trouver une définition de l'emphase. Elle n'est que très rarement définie dans la littérature alors que tout le monde en parle. Aussi avons-nous pensé qu'il était nécessaire de clarifier cette notion, de tenter d'en donner une définition et d'en extraire les corrélats acoustiques et prosodiques.

Nous sommes cependant confrontée au problème de la méthode. Devons-nous nous inscrire dans une théorie bien particulière ? Devons-nous nous contenter d'analyses phonologiques ou phonétiques ?

On trouve ce genre de débat dans la littérature et ce n'est pas un problème simple. t'Hart, Collier & Cohen (1990) résument bien la chose. Dans l'idéal, une théorie de l'intonation devrait comprendre une composante linguistique et une composante phonétique. La partie phonétique devrait s'intéresser aux aspects physiologiques, acoustiques et perceptuels de l'intonation et en élucider les fonctionnements. La partie linguistique devrait donner une interprétation phonologique des faits acoustiques et une explication pragmatique des fonctionnements de l'intonation dans le cadre de l'interaction communicative. Enfin, la théorie devrait donner des précisions sur le fait que la performance mélodique du locuteur résulte de l'interaction entre son intention communicative et les moyens périphériques de ses appareils vocal et perceptif.

Ainsi, deux options se présentent au chercheur : soit il choisit une approche linguistique, mais risque de manquer certains phénomènes acoustiques, soit il adopte une approche instrumentale et risque de passer à côté des essentiels de la communication :

The student of intonation faces a dilemma : either he chooses the linguistic approach at the risk of overlooking phonetically important features, or he opts for an instrumental-phonetic angle, thus increasing the chance of missing the communicatively relevant essentials. (t'Hart, Collier & Cohen, 1990 : 2)

C'est face à ce dilemme que nous nous sommes trouvée et nous avons choisi d'opter pour la deuxième solution. En effet, une étude expérimentale nous permettra de fonder nos affirmations sur du concret et de l'empirique. Nous essaierons néanmoins de ne pas oublier les mises en garde de t'Hart, Collier & Cohen en suivant les conseils de Liberman et Pierrehumbert (1984 : 232) :

We feel that hybrid strategies, in which the methods of phonology and phonetics are simultaneously exercised, will play a crucial role in improving theories of human speech and language.

Nous allons tenter dans ce travail de définir quels paramètres permettent de reconnaître l'emphase dans le discours. Quelques chercheurs se sont penchés sur cette problématique, mais toujours, à l'exception de Selting (1994) et de Uhmann (1992), en se basant sur de la parole de laboratoire. L'emphase en situation de communication

réelle a été largement négligée (cf. Beckman, 1997). C'est pourquoi nous avons souhaité fonder notre étude sur de la parole naturelle. Ces postulats de base étant définis, nous avons divisé notre travail en cinq parties.

Nous faisons tout d'abord un état de la question : même si l'emphase n'est pas un sujet qui a été amplement étudié, l'état de l'art est conséquent car la notion d'emphase est liée à beaucoup d'autres concepts que nous récapitulons donc dans la première partie de notre étude afin d'y voir un peu plus clair.

La deuxième section traite de la base de données que nous avons choisie et des méthodes que nous avons employées pour notre recherche. Nous avons déjà dit que l'emphase, et la parole en général, en situation réelle de communication avait été largement négligée, et que nous souhaitions donc analyser du discours spontané. Mais encore fallait-il choisir une base de données. D'autre part, la méthode d'analyse est essentielle : les études antérieures sur l'emphase ont pour point de départ un préconçu qui détermine les passages ou mots emphatiques. Nous avons souhaité nous démarquer en ce qui concerne la méthode et notre travail est fondé sur une étude perceptive qui nous permet de dégager l'emphase dans notre base de données. Nous détaillons cette méthode et en discutons largement dans le chapitre 2.

Pour procéder à l'analyse acoustique de notre base de données, nous avons utilisé le logiciel PRAAT (Boersma & Weenik, 1996). Il s'agit d'un logiciel récent très performant, et de plus gratuit et multi-plateforme. Nous avons utilisé tout au long de notre étude des outils à la pointe dans leur domaine, afin que notre travail s'inscrive dans le cadre des technologies modernes disponibles au chercheur. Le chapitre 3 relate, décrit et explique les analyses acoustiques que nous avons choisi d'entreprendre et qui nous permettent de déterminer des variables pertinentes en relation avec l'emphase.

Dans la quatrième section, nous décrivons l'analyse statistique des paramètres acoustiques et prosodiques décrits dans le chapitre précédent. Nous avons là encore opté pour un logiciel tout récent à la pointe de la recherche, appelé CRUISE (Kim & Loh, sous presse). Cette analyse objective nous permet d'extraire les paramètres les plus révélateurs quant à l'emphase.

Enfin, la cinquième section est à nouveau largement expérimentale mais traite cette fois d'une étude réalisée sur des signaux manipulés grâce au synthétiseur de parole MBROLA (Dutoit *et al.*, 1996 ; Dutoit, 1997). Le but de cette nouvelle expérience est

également de tenter de définir quels sont les paramètres les plus pertinents de l'emphase, mais à partir d'une étude perceptive, subjective, fondée sur la perception humaine, et non d'une étude statistique objective.

Nous achevons ce travail par une discussion des résultats et nous nous penchons sur des perspectives de recherches futures, en particulier en ce qui concerne un élargissement des paramètres, non plus seulement acoustiques, mais sémantiques, syntaxiques, pragmatiques, et également sur une perspective d'utilisation de ces paramètres pour une amélioration de la parole de synthèse.

1 Chapitre 1

*L'emphase : définitions et études
antérieures*

Avant de commencer tout travail sur l'emphase, il nous faut donner un aperçu de ce qui a été dit sur ce sujet, et des travaux qui ont été entrepris dans ce domaine. C'est ce que nous allons tenter de faire dans ce chapitre sans pour autant prétendre être exhaustive sur tous les points.

Les descriptions et études de l'emphase que l'on trouve dans la littérature présentent un flou théorique considérable. On peut parler d'une limitation de cette notion qui n'est définie, comme nous allons le voir, que par très peu de linguistes. L'emphase est souvent associée à d'autres termes et nous essaierons de clarifier toutes les notions qui entrent en jeu.

Il est plus souvent question d'emphase dans des études expérimentales réalisées par des chercheurs sur des domaines parfois plus larges, et nous résumerons les travaux qui nous semblent les plus importants dans la deuxième section de ce chapitre.

Ainsi nous verrons que certains linguistes ont associé l'emphase à l'émotion, l'emphase étant une façon de rendre une émotion, et des études expérimentales ont été effectuées dans ce cadre, tentant d'établir des relations entre les paramètres acoustico-prosodiques et les émotions exprimées par le locuteur. Nous verrons quelles en sont les limites.

De nombreux travaux empiriques traitent de ce qu'on appelle en anglais « stress », « accent » et « prominence », notions que nous aurons définies dans la première section, et nous verrons de quelle façon elles peuvent être utiles à l'étude de l'emphase.

Des études plus précisément ciblées sur l'emphase ont été réalisées récemment et nous verrons quels paramètres ont été étudiés. Des études pragmatico-acoustiques en particulier semblent bien définir le phénomène.

1.1 Définitions

Il est nécessaire avant toute chose de se demander de quelle façon l'emphase est définie en tant que terme linguistique. Nous sommes frappée en effet par le fait que les mots « emphase » ou « emphatique » apparaissent assez souvent dans différents ouvrages sans que le terme ne soit jamais défini au préalable. Ainsi il semble que ce terme soit tenu pour acquis par tous. C'est la raison pour laquelle nous nous sommes demandée ce que voulait dire « emphase » pour la plupart des auteurs.

1.1.1 Définitions générales

Nous avons tout d'abord procédé à une recherche toute simple de la définition que l'on pouvait trouver du terme « emphase » dans les différents dictionnaires et encyclopédies de linguistique.

C'est peut-être justement parce qu'il est difficile de donner une définition de l'emphase que beaucoup de linguistes emploient ce mot sans le définir. Certains ne le mentionnent d'ailleurs pas du tout (Bright, 1992 ; Collinge, 1990 ; Asher, 1994) ou en tout cas pas dans son sens linguistique mais rhétorique (Bussman *et al.*, 1996 : 145) :

Emphasis : (Greek emphasis 'exposition', from emphaínein 'to exhibit', 'to indicate'). Also known as 'significatio', emphasis means to imply more than is actually stated. This can be accomplished by choosing an exceptionally strong word or phrase : *Be a man !* Emphasis can also be achieved by saying less than you mean, implying more than you say : for example : *He has such charm ...* Various devices can create emphasis : tautology, pleonasm, cliché, simile, litotes, interjection, exclamation.

Cette définition plutôt axée sur les figures de style est néanmoins intéressante quant à sa dimension sémantique. L'idée d'implication nous semble pertinente et nous y reviendrons¹.

¹ Cf. 1.5.

Ceux qui définissent le terme dans son sens plus précisément linguistique l'associent forcément à d'autres notions ou concepts qu'il nous faudra étudier.

Les grammairiens ou linguistes traditionalistes et les pragmaticiens parlent de mise en relief ou d'insistance :

le langage tend constamment à mettre en relief ce qui est nécessaire (principe d'emphase) (Mounin, 1974 : 123)

Le terme d'emphase - emprunté à la rhétorique - désigne les phénomènes d'insistance qui peuvent affecter les constituants de la phrase. (Arrivé *et al.*, 1995 : 244)

la mise en valeur d'une expression, appelée aussi emphase, insistance... (Ducrot & Todorov, 1972 : 346).

Il est intéressant de noter qu'un syntacticien transformationaliste comme Dubois (1973 : 186) utilise le mot « accent » (dans le sens de mise en relief) en plus du terme « topique » :

Le terme d'emphase, repris à la rhétorique, désigne en syntaxe transformationnelle un accent particulier porté sur un constituant de la phrase. Le constituant emphase topicalise le syntagme porteur d'emphase (celui-ci devient le topique de la phrase) (on appelle topique le sujet du discours défini comme 'ce dont on dit quelque chose').

Les énonciativistes (comme Groussier & Rivière, 1996) associent quant à eux emphase et focus :

en linguistique, emphase = focalisation (p. 66) ; focalisation = mise en relief (p. 84).

Certains (Trask, 1993, par exemple) mêlent « focus », « old and new information », « contrast » :

emphasis : a very general term for any phenomenon which serves to draw particular attention to some element in a sentence or utterance, either to put that element in focus or to contrast it with some other element. (p.89)

focus : special prominence given to some element which represents the most important new information in that sentence or which is explicitly contrasted with something else. (p.105).

Enfin les linguistes plus proches de la phonologie exposent en tout premier l'aspect prosodique :

Emphasis : high pitch or heavier stress or both, given to a word or part of a word to indicate that it has special importance in an utterance (Hartmann & Stork, 1972 : 76).

Nous voyons qu'il va donc être difficile de réaliser un travail sur l'emphase sans procéder à un tri de toutes ces notions et de tous ces concepts.

Mais qu'ils le mentionnent en premier lieu ou non, tous s'accordent à dire qu'un des moyens d'exprimer l'emphase est la prosodie et c'est cela qui nous semble, bien entendu, le plus pertinent :

Le signifié emphase peut se trouver représenté uniquement au niveau phonologique (intonation particulière frappant un mot de la phrase), mais aussi par des constructions syntaxiques. (Dubois, 1973 : 186)

Une configuration prosodique spécifique est le marqueur obligatoire de l'opération de focalisation². Il peut fonctionner seul ou en association avec d'autres. (Groussier, Rivière, 1996 : 169)

Emphasis in spoken English is often achieved merely by stressing the emphasized element, but English and other languages also exhibit a range of grammatical means for expressing emphasis such as particles, distinctive word order, and clefted constructions. (Trask, 1993 : 89)

L'emphase peut se manifester uniquement par des traits prosodiques. Elle peut entraîner en outre des modifications de la phrase. (Arrivé, 1995 : 244)

Emphase : phonologique : principe d'emphase : Terme dû au phonéticien Paul Passy qui rendait responsables des changements phonétiques deux principes antagonistes : [...] loi du moindre effort, [...] principe d'emphase. (Mounin, 1974 : 123).

Nous n'avons pas encore mentionné ici l'ouvrage encyclopédique de Crystal (1993) car il ne donne pas de définition du terme emphase, mais il parle d'émotion et plus particulièrement de « the emotional function of intonation ». Il nous paraît impossible de séparer l'emphase de l'émotion car l'emphase est intimement liée au côté émotionnel de la langue. Parmi les réactions émotionnelles que Crystal mentionne, toutes ne feront

² Nous avons vu ci-dessus que pour Groussier et Rivière, focalisation = emphase.

probablement pas appel à l'emphase, mais si emphase il y a, elle sera forcément, nous semble-t-il, incluse dans ces réactions.

The most obvious function of intonation is to express a wide range of attitudinal meanings – excitement, boredom, surprise, friendliness, reserve and many hundreds more. Here, intonation works along with other prosodic and paralinguistic features to provide the basis of all kinds of vocal emotional expression. (Crystal, 1993 : 171).

Crystal mentionne aussi le côté visuel de la langue relié à l'emphase, point que nous ne pourrions passer sous silence :

Gestures and head movements tend to coincide with points of emphasis. (Crystal, 1993 : 402).

Ces considérations sur l'émotion nous amènent à citer d'autres linguistes qui relient directement l'emphase et l'émotion, ou ce qu'ils nomment aussi attitudes (« attitudes »), ou encore implication (« involvement »), à commencer par Quirk *et al.* (1985 : 1414) qui distinguent deux types d'emphase :

emphasis given by information focusing
et
purely emotive emphasis.

Pour Selting (1994 : 376),

Signals of 'emphasis' or 'surprise' are important and common emotive displays in which 'more-than-normal' involvement is expressed linguistically by cues that suggest the appropriate interpretive frames.

Cette idée d'implication plus importante qu'à la normale nous semble tout à fait pertinente et rejoint d'ailleurs la définition rhétorique de Bussman *et al.* (ci-dessus).

L'emphase étant en fait le résultat d'une émotion, il est impossible de ne pas en étudier l'intonation, et Bolinger (1986 : 195) a été l'un des premiers à le clamer et à le répéter :

Intonation is part of a gestural complex whose primitive and still surviving function is the signalling of emotion.

Suivant Classe (1939) qu'il cite d'ailleurs, Bolinger (1989 : 68) va même jusqu'à dire que tout discours exprime forcément une émotion :

If there is neutral speech, it is not speech in which emotion has been eliminated but tamed.

En résumé nous pouvons dire que la notion d'emphase est associée aux termes focus, topicalisation, contraste, proéminence et plus largement émotion. Nos recherches se pencheront donc aussi sur ces notions.

Enfin ce petit survol des différentes définitions que chacun donne de l'emphase nous permet d'affirmer qu'il s'agit d'une notion qui s'étend sur plusieurs domaines :

- la pragmatique (« old and new information », idée d'implication de la part du locuteur) ;
- la sémantique (contraste, topique) ;
- la syntaxe (transformations syntaxiques) ;
- la prosodie ;
- et même la gestuelle.

Le domaine de la prosodie reste cependant celui que tout le monde s'accorde à donner comme indispensable, c'est la raison pour laquelle nous nous y attacherons principalement, sans pour autant négliger complètement les autres mais en les gardant au contraire toujours à l'esprit.

1.1.2 Focus, topique, et information nouvelle

Gussenhoven (1983 : 14) résume bien les choses en expliquant que les notions de focus, topicalisation ou information nouvelle se recoupent dans la littérature selon les différents auteurs :

- « focus », chez par exemple Chomsky (1969), Jackendoff (1972), Quirk *et al.* (1972) ou Ladd (1980) ;
 - « comment », chez Bloomfield (1933) ou Schmerling (1976) ;
 - « rheme », chez les linguistes de l'école de Prague ;
 - « new information », chez Halliday (1967b) ou Chafe (1970 ;1976)
- ont le même sens.

Respectivement, leurs opposés sont :

- « presupposition » : Chomsky (1969), Jackendoff (1972), Quirk *et al.* (1972) ; ou « deaccenting » : Jackendoff (1972), Ladd (1980) ;
- « topic » : Schmerling (1976) ;
- « theme » : école de Prague ;
- « given » ou « old information » : Halliday (1967b), Chafe (1970 ;1976).

Ces notions ne sont pourtant pas équivalentes pour tous.

Quirk *et al.* (1985 : 1362) opposent thème et focus, le thème étant ce qui vient en premier dans un énoncé - définition empruntée aux linguistes de l'école de Prague (cf. par exemple Firbas, 1972) et indiquent qu'ils appellent parfois focus rhème. En effet, selon eux, le thème vient en premier dans une phrase, il s'agit donc du contexte qui est en général déjà connu (information ancienne) et donc inaccentué, alors que ce qui vient après le contexte est de l'information nouvelle, qui constitue le « focus » du message et qui est logiquement proéminent au niveau prosodique (cf. p. 1357). Ainsi Quirk *et al.* mélangent thème, information ancienne et rhème, focus et proéminence accentuelle ou prosodique. Ils distinguent cependant quand même l'opposition information ancienne / information nouvelle, qui est établie par le contexte et est donc extralinguistique, et l'opposition thème / focus qui est définie linguistiquement en termes de position et prosodie respectivement. Le thème n'est donc normalement pas emphatique d'après leur définition (cf. p. 1361).

Bouscaren *et al.* (1992 : 12n) mettent d'ailleurs en garde le lecteur afin qu'il ne confonde pas thématisation, à savoir choisir un élément comme terme de départ d'un énoncé, et emphase ou focus. On note que là encore emphase et focus sont sur le même plan alors que thème et emphase ne le sont jamais.

Enfin Peeters (1999) va dans le même sens en donnant les définitions suivantes des deux termes focalisation et thématisation :

[La focalisation] constitue une forme marquée de la construction taxinomique du sens, car elle rend pratiquement explicite ce qui est implicite dans tout sens, à savoir sa nature différentielle. La focalisation est donc un procédé sémantique qui, suivant les situations, peut prendre toutes sortes de valeurs comme l'insistance, la polémique, l'emphase, la précision, etc... (p. 48)

alors que

[La] thématisation consiste à marquer, non pas qu'on a choisi taxinomiquement un sens plutôt qu'un autre, mais que la phrase suit un schéma organisationnel (sujet + prédicat + objet + etc.) et que le sens global de l'énoncé se fragmente en circonstants qui ne vont pas l'un sans l'autre. [...] La thématisation consiste à marquer à partir de quoi s'organise ou progresse l'énoncé. (p. 53)

Ces trois définitions de la thématisation (Quirk *et al.*, Bouscaren *et al.*, et Peeters) prouvent que cette notion n'est absolument pas liée à l'emphase et par conséquent nous nous en desintéresserons.

Lambrecht (1994) tente de définir et de distinguer « topic » et « focus ». Pour lui aussi, le thème est ce qui vient en premier dans une phrase et ce n'est pas nécessairement ce dont on parle, selon la construction syntaxique choisie, et il oppose thème à topique, le topique étant la chose à propos de laquelle on parle :

Topic : a referent is interpreted as the topic of a proposition if in a given situation the proposition is construed as being about this referent, i.e. as expressing information which is relevant to and which increases the addressee's knowledge of this referent. (Lambrecht, 1994 : 131)

Le topique est une catégorie pragmatique selon Lambrecht alors que la notion de focus est une composante sémantique. Le focus est ce qui est imprédictible, ce qu'on ne peut pas retrouver au moment du discours. Lambrecht le définit comme suit :

The semantic component of a pragmatically structured proposition whereby the assertion differs from the presupposition. (p. 213)

Là encore il nous semble que la notion de topique ou topicalisation n'est pas directement liée à l'emphase (contrairement à celle de focus) et nous n'y reviendrons pas.

Hirst & Di Cristo (1998) ne partagent pas non plus tout à fait l'avis de Gussenhoven, selon lequel tous ces concepts ont le même sens sous des noms différents, variant en fonction des auteurs et des écoles linguistiques. Même s'ils commencent par présenter les notions de « focus/presupposition » équivalentes à celles de « theme/rheme, given/new or topic/comment », ils précisent ensuite que « focus » est une notion plus générale que les autres qui recouvre à la fois les notions d'information nouvelle et ancienne et celles d'emphase et contraste :

[a notion] which attempts to account by means of a single mechanism both for the given/new aspects of an utterance and at the same time for what has traditionally been treated under a separate heading as emphasis and contrast. (p.31)

De nombreux autres linguistes associent focus et emphase.

Pour Halliday (1967b : 204) l'information mise en « focus » est une sorte d'emphase :

information focus is a kind of emphasis, that whereby the speaker marks out a part of a message block as that which he wishes to be interpreted as informative.

Ladd (1980 : 213) va encore plus loin en associant focus, intonation et emphase :

focus is signaled solely by the location of the accent ; various intonational characteristics such as greater volume and widened pitch range can also be used to signal what might be called emphasis.

La notion de focus est donc quant à elle essentielle en relation avec l'emphase et il est nécessaire de l'approfondir. Plusieurs linguistes distinguent le focus étroit (« narrow focus »), et le focus large (« broad focus ») et même parfois le focus neutre (« neutral focus ») et les notions d'information nouvelle et ancienne sont sous-jacentes. Il nous faut donc éclaircir ces points. Nous commencerons par un rappel historique rapide.

1.1.2.1 Focus, présupposition et information

C'est Chomsky (1969) qui parle le premier de l'opposition entre focus et présupposition. Le focus d'une phrase est cette partie de la phrase qui, quand on la remplace par une variable x, livre la partie présupposée de la phrase. L'accent tombe sur le mot focalisé. Mais le focus peut s'étendre au-delà du mot accentué.

Focus est donc lié à information, comme l'explique Jackendoff (1972). Le focus est l'information que le locuteur et l'interlocuteur ne sont pas censés connaître tous les deux :

the information in the sentence that is assumed by the speaker not to be shared by him and the hearer. (Jackendoff, 1972 : 230)

Quant à la présupposition, il s'agit de l'information partagée par les deux :

the information that is assumed by the speaker to be shared by him and the hearer. (*ibid.*)

Bolinger (1972) lie également focus et information. Il appelle les éléments qui sont les plus inattendus, avec une faible probabilité, « points of information focus ». Selon lui, ce sont ces éléments-là qui sont accentués dans la phrase et c'est leur poids sémantique qui détermine le placement de l'accent. Dans un exemple comme :

I have a POINT to make vs. *I have a point to Emphasize,*

le verbe « emphasize » est plus inattendu, moins prédictible que « make » dans cette expression, il a donc plus de poids sémantique, c'est pourquoi il est accentué.

Bolinger ajoute quand même que le contexte aide à définir où le « point of information focus » va se trouver.

Schmerling (1976) critique ce critère de prédictabilité grâce aux deux exemples bien connus :

(i) *JOHNson died.*

(ii) *Truman DIED.*

Comme le président Johnson était en bonne santé lorsqu'il est mort d'une crise cardiaque alors que l'on s'attendait à la disparition de Truman qui était mal en point depuis quelque temps, Schmerling avance l'argument que selon les critères de prédictabilité de Bolinger on aurait dû avoir les schémas accentuels inverses : l'accent sur « died » en (i) et sur « TRUman » en (ii). Schmerling parle donc de « news sentences » comme en (i) où l'on parle d'un événement ou d'un état de choses, et de « topic-comment sentences », comme en (ii) qu'elle décrit comme étant des phrases à propos du sujet (grammatical) de la phrase. Dans les « news sentences », le verbe a un accent moindre que le sujet ou l'objet. Dans les « topic-comment sentences », à la fois le topique (« topic ») - ce dont on parle - et ce qu'on en dit - « comment » - sont accentués, avec un accent plus fort sur l'élément le plus à droite.

Cela revient à parler d'information nouvelle et ancienne. L'accent nucléaire signale le focus et ce qui est focus est de l'information nouvelle.

Couper-Kuhlen (1986 : 44f) démontre pourtant que lorsque l'on met les phrases en contexte et que l'on étudie ce qui se passe dans le discours, on trouve de nombreux cas où de l'information non nouvelle est mise en relief : le placement de l'accent dépend

alors entièrement des motivations et des choix du locuteur. C'est toute la dimension sociale de l'accent selon elle, et l'on touche déjà là au côté pragmatique de l'accentuation et donc de l'emphase, sur lequel nous reviendrons.

Cruttenden (1986 : 88) pense aussi que le focus ne correspond pas toujours à l'information nouvelle :

... in many cases (though not all)³ the focal part of an intonation-group can be said to involve new information

et rappelle que le fait de ne pas avoir d'accent nucléaire ne veut pas nécessairement dire information ancienne. D'autre part, Cruttenden mentionne trois cas dans lesquels de l'information ancienne fait partie du focus :

si l'information ancienne est contrastive :

All three of them had a GO / Only the MOTHER / was sucCESSful (p. 91)

« Mother » est le focus alors qu'on y a fait référence auparavant en disant « all three of them ».

s'il s'agit d'un écho (questions reprenant ce qui vient d'être dit et impliquant le plus souvent une note d'incrédulité ; reprises exclamatives ; questions pour faire répéter) :

(I didn't go after all.) You didn't GO ?

(He's got a distinction) A DisTINction !

(We went to Manchester this morning.) What did you go to MANchester for ?
(p.92)

si le locuteur n'est pas d'accord avec ce qui était présupposé dans l'énoncé du locuteur précédent. C'est ce que Cruttenden nomme « counterpresuppositionals » qui constituent une sous-catégorie de « insists » :

(What sort of b.o. powder do you use ?) I don't USE b.o. powder. (p. 93)

Ces cas attirent tout particulièrement notre attention car nous pensons que ce sont des exemples que l'on pourrait qualifier d'emphatiques au sens le plus intuitif du terme. Cruttenden parle de focus étroit, car à la suite de Ladd (1980), il distingue deux types de

³ C'est nous qui soulignons.

focus, qui sont essentiels en ce qui concerne la portée du focus : focus large et focus étroit.

1.1.2.2 Focus large et focus étroit

Comme Chomsky (1969) le précisait déjà, le focus peut s'étendre au-delà du mot accentué.

Pour Cruttenden (1986 : 81) et bien d'autres, le moyen de mettre en relief est de placer le nucleus sur un élément :

Nucleus placement is the principal means of focusing in English.

On parle de focus large (« broad focus ») lorsqu'un groupe intonatif tout entier est mis en relief :

the whole of the intonation group is in focus. (p. 81)

Et on parle de focus étroit (« narrow focus ») lorsqu'un seul élément du groupe intonatif est mis en relief :

a grammatical constituent which forms part of the intonation unit is brought into focus. (p. 81).

Quand il y a focus large, l'énoncé est considéré comme « all-new » (tout nouveau) ou « out-of-the-blue » (inattendu) ou pourrait être prononcé en réponse à la question : « Que s'est-il passé ? » (p. 81). On peut dire qu'on a un focus large quand rien n'est particulièrement emphatique justement, c'est pourquoi on le nomme aussi « normal stress » et c'est le dernier élément lexical qui porte l'accent nucléaire. Ce concept de « normal stress » est amplement discuté dans la littérature, en particulier par Schmerling (1976) qui le rejette et préfère parler de focus non marqué (« unmarked focus ») ou Ladd (1980 : 76) qui réconcilie l'approche sémantique de Bolinger (cf. 1.1.2.1., les « points of information focus ») et les approches syntaxiques de Chomsky et Halle (1968 : la NSR – « nuclear stress rule ») en expliquant qu'un contour intonatif normal (« normal stress ») est la façon de prononcer un focus non marqué. Ainsi Ladd (1980 : 76) reformule les deux théories :

Bolinger's formulation might be changed from 'accent goes on the point of information focus' to 'accent goes on the point of information focus, unless

the focus is unmarked, in which case the accent goes on a location determined by the syntax.'

Le focus reste un phénomène sémantique auquel on a simplement ajouté un cas non marqué.

we can reinterpret the Trager-Smith-Chomsky-Halle normal stress rule in terms of focus. Instead of saying that normal stress goes on the rightmost accentable item in the sentence, we will say that accent - in general - goes on the rightmost accentable item of the 'focus constituent'. If the focus constituent is the whole sentence, we get 'normal stress' ; if not, we get a narrow focus on the constituent identified by the placement of accent. (p.77)

Ceci veut dire que dès que l'on a un accent nucléaire sur un élément grammatical ou bien sur un élément lexical non final, on parle de focus étroit.

Cruttenden (1986) donne quand même quelques cas où le focus est large sans que l'on ait pour autant l'accent nucléaire sur le dernier élément lexical :

event sentences, final adverbials, adjectival wh- objects (cf. p. 83)

Cruttenden attire donc notre attention sur des problèmes de syntaxe et plus particulièrement de classes syntaxiques. En effet, le fait même de distinguer les éléments lexicaux des éléments grammaticaux, comme il est courant de le faire (tous s'accordent à dire que les éléments grammaticaux ne sont normalement - par « normal » on entend en contextes non contrastifs ou non emphatiques - pas accentuables) illustre ce point et Cruttenden ajoute que parmi les éléments lexicaux, les noms sont plus à même de se trouver sous l'accent que les autres classes de mots :

within the different classes of lexical items, nouns are certainly more accentable than the other classes. (Cruttenden, 1986 : 87)

Ceci explique largement selon Cruttenden pourquoi l'accent nucléaire ne se trouve pas sur le dernier élément lexical dans les trois catégories dissidentes mentionnées plus haut.

Quand il s'agit de focus étroit, une partie de la phrase est hors de la portée du focus pour différentes raisons (parce qu'elle est redondante par exemple). La partie restante qui se trouve focalisée contient comme pour le focus large l'accent nucléaire sur l'élément lexical final.

Le problème que cela soulève est que la place de l'accent nucléaire seule ne suffit pas à indiquer quelle est la portée du focus. Un accent prénucléaire peut parfois être déterminant :

Sometimes the beginning of the extent of focus is indicated by a prenuclear pitch accent [...] but this is not obligatory. (Cruttenden, 1986 : 88)

Cependant, cela n'est qu'un pseudo-problème selon Cruttenden puisque le contexte permet d'éclaircir les choses. C'est à ce moment-là que rentrent en jeu les concepts d'information nouvelle et information ancienne (cf. 1.1.2.1.).

Ladd (1980 : 81ff) parle quant à lui de « default accent » pour élucider le même problème (le fait que dans le focus large, l'accent nucléaire ne soit pas toujours sur le dernier élément lexical), ce qui rejoint selon lui « deaccenting ». En effet lorsque l'on choisit de désaccentuer certains éléments, d'autres se retrouvent accentués par défaut. Mais cela n'implique pas que l'accent devienne alors forcément contrastif. S'il est contrastif, il s'agit d'un focus étroit.

Ladd précise en note (pp. 213-4) qu'il sépare les aspects accentuels et intonationnels : le focus étroit est signalé uniquement par le lieu de l'accent⁴. D'autres paramètres tels que l'intensité et la hauteur mélodique (« widened pitch range ») peuvent aussi être utilisés pour signaler l'emphase. Il est possible d'avoir un focus étroit sans paramètres intonationnels d'emphase, et il est de la même façon possible d'avoir les paramètres intonationnels de l'emphase sans avoir un focus étroit. Cela est néanmoins rare, rajoute Ladd. Ici nous aurions bien sûr apprécié que Ladd soit plus prolixe car il ne donne qu'un exemple où l'on a de l'emphase et pas de focus étroit et n'explique rien :

Wha wor that
t in the ld are you doing for ?

Hirst (1998) rapproche les termes focalisation et emphase, qui sont tous deux des manières de mettre en relief une partie de l'énoncé, en expliquant que le terme emphase est l'équivalent de focus étroit :

[T]he term focus, more specifically narrow focus, has in a number of recent works replaced the more traditional term of emphasis to refer to the way in which a speaker gives an optional prosodic highlighting to part of an utterance. (p. 66)

⁴ Il faut entendre ici l'accent nucléaire.

Cependant une nuance existe tout de même et Hirst préfère utiliser le mot « emphasis » à celui de « focus » car on ne peut parler de « narrow » ou « broad » focus sur un seul mot, « focus » étant une notion syntagmatique, alors que l'on peut parler d'emphase, qui est une notion paradigmatique :

[W]hereas emphasis is basically a paradigmatic notion, *i.e.* a given item is either emphatic or non-emphatic, focus is basically syntagmatic : focus is on one element of a sequence (more strictly it applies to one node of a tree). This implies that while it is possible to refer to emphatic or non-emphatic versions of an utterance consisting of a single word (such as *no* vs. *NO* !), it is not possible within standard treatments of focus to distinguish broad and narrow focus on a single item. (Hirst & Di Cristo, 1998 : 31-2)⁵

Nous avons vu que dans la plupart des cas, seule l'information nouvelle peut normalement se trouver dans la portée du focus, pas l'information ancienne sauf si celle-ci est par exemple contrastive (cf. Cruttenden, 1986). Le contraste rejoint alors le focus étroit (cf. par exemple Couper-Kuhlen, 1986 : 48) et donc l'emphase. Cela nous amène de ce fait à nous interroger sur la notion de contraste qui est souvent mentionnée elle aussi en relation avec celle d'emphase.

1.1.3 Contraste

Cruttenden (1986) parle de contrastivité et donne une définition de « contrastif » : une comparaison avec un ensemble limité (p. 90). Tout le problème est de savoir ce que l'on entend par ensemble limité. Cruttenden insiste bien sur la difficulté d'étudier ce type de notions car elles ne sont définies clairement par personne et semblent acquises par tous :

[...] like many other notions we are dealing with elusive concepts which seem to exist in speakers' and listeners' minds but which defy stringent definition.
(p. 90)

Brown *et al.* (1980 : 136) différencient l'information contrastive et l'information nouvelle, la première ayant « a higher pitch », et ils ajoutent :

⁵ Cf. Hirst (1998 : 66) pour plus de détails sur ce point.

there also appears items raised in the speaker's pitch range which can be interpreted either as in non specific contrast or, more straightforwardly, as being emphasized by the speaker.

Pour Couper-Kuhlen (1986), il existe peu de différence entre emphase et contraste. Elle pense que cela dépend des linguistes et qu'il s'agit souvent d'une question de vocabulaire. Selon elle, un schéma intonatif contrastif est un schéma qui diffère du schéma normal⁶ : « deviation from normal pattern » (p. 47). Néanmoins, un accent contrastif peut tomber en position normale, et de plus, un accent peut ne pas être à l'endroit normal sans être pour autant contrastif. C'est pourquoi elle emploie le terme « emphatic accent placement ». Couper-Kuhlen croit que le contraste sémantique strict est identifiable car il s'accompagne d'une configuration mélodique distinctive, et à la suite de Brown *et al.* (1980), elle dit qu'un contraste non spécifique est une emphase⁷. Elle ajoute qu'à la fois le contraste spécifique et le contraste non-spécifique impliquent une présupposition, et se démarquent en cela de l'information nouvelle, et que :

contrast/emphasis are distinct from new information, distinctive pitch configurations corresponding to these two categories. (p. 137)

Cette dernière citation montre bien que Couper-Kuhlen associe les deux termes contraste et emphase. Elle ne les tient cependant pas pour équivalents car un contraste peut ne pas être emphatique et une emphase peut ne pas être contrastive.

Hirst (1977b) distingue deux types d'emphase à la suite de Coleman (1914)⁸ : « contrast emphasis » et « intensity emphasis ». Il précise (Hirst & Di Cristo, 1998 : 31) que traditionnellement, contraste et emphase ont toujours été associés, le premier étant inclus dans le second :

Contrast has traditionally been held as a subset of emphasis,

mais ne partage pas entièrement cette opinion. Si l'on prend la phrase

⁶ Pour Couper-Kuhlen (1986), le schéma normal est celui où le dernier élément accentuable (« lexical item ») prend l'accent nucléaire.

⁷ Un contraste spécifique signifie que le locuteur choisit quelque chose et cela implique simultanément que toutes les autres choses possibles ne sont pas les bonnes. Dans un contraste non spécifique, le locuteur présuppose que quelque chose avait peu de chance d'être choisi mais il choisit ce quelque chose là quand même.

⁸ Nous reviendrons sur l'étude de Coleman plus en détails : cf. 1.2.2.1.

I do like your tie

Hirst argumente que l'intonation va changer selon que l'on veut exprimer une emphase contrastive ou une emphase d'intensité. Dans le premier cas, la tonique de tête (« I » ici) sera moins haute (plutôt dans le médium) que dans le deuxième cas où elle sera plus probablement dans l'aigu. En effet, l'emphase

ne se manifeste que par le comportement tonal de l'ensemble de l'unité intonative dans laquelle ce morphème⁹ est réalisé (Hirst, 1987 : 244).

C'est grâce à cela, grâce au « high pre-head » selon Hirst (1987), que l'on peut distinguer une phrase emphatique avec l'élément mis en relief en fin de phrase d'une phrase neutre. Ces réflexions nous semblent tout à fait essentielles et sont des indices qui nous permettront de déterminer quels paramètres étudier.

Bolinger (1961) parle lui aussi de contraste pour des mots ou des syllabes mises en relief (« phonetic highlighting »). Il distingue « contrastive stress » et « contrastive accent » de façon très intéressante car, nous semble-t-il, ces deux concepts sont des façons de rendre une emphase. Ce que Bolinger appelle « contrastive accent » est le phénomène grâce auquel :

two or more items are counterbalanced and a preference indicated for some member or members of the group. (p. 83)

Bolinger démontre que « contrastive accent » n'est pas définissable phonétiquement grâce au contour intonatif comme le pensent la plupart des linguistes de l'époque. Il donne entre autres l'exemple de la phrase :

My mother is coming.

Traditionnellement, on dit que l'accent le plus neutre se trouve sur « coming » et tout autre accent serait contrastif : « my » serait en contraste avec « your » ou « his » par exemple, « mother » avec quelqu'un d'autre, « is » avec « is not ». Or si l'on répond :

My mother is coming

avec l'accent sur « mother » en réponse à la question :

Why are you in such a hurry to get home ?

⁹ Il s'agit du morphème porteur d'emphase.

la réponse n'est pas du tout contrastive. C'est pourquoi il pense qu'on ne peut différencier phonétiquement un accent contrastif d'un autre accent :

the semantic peak includes contrastive accents along with other accents ; and there is no predictable phonetic difference. (p. 84).

Néanmoins, on peut argumenter que tout est contrastif. La suggestion tout à fait innocente « Let's have a picnic » peut être en contraste avec faire n'importe quoi d'autre que pique-niquer (cf. p. 87).

La notion de « contrastive accent » est en fait pour Bolinger plutôt abstraite, pragmatique ou sémantique bien plus que phonétique. C'est pour cela qu'il distingue « accent » et « stress ».

Sa définition de « stress » par contre est concrète et phonétique. « Stress » est en quelque sorte la réalisation de « accent » :

The stressed syllable of a word is the one where the accent falls if there is an accent, which is to say that it is the syllable with a potential for accent (p. 88).

S'il s'agit de « contrastive stress » (dû à « contrastive accent »), le mot sera prononcé différemment (allongement d'une voyelle inaccentuée, explosion anormalement appuyée d'une consonne, petite déformation d'une voyelle pour rapprocher de la graphie, ou encore le « stress » se déplace sur une syllabe qui ne porte pas normalement le « stress » : « stress shift »), ou bien un mot qui est habituellement « unstressed » pourra porter un « stress » (p. 95). On peut trouver des mots sur lesquels s'appliquent un « contrastive accent » et un « normal stress » (et c'est le cas le plus courant) ou bien des mots avec à la fois « contrastive accent » et « contrastive stress ». S'il y a « contrastive stress », il y a forcément « contrastive accent ». L'inverse n'est pas vrai.

1.1.4 « accent », « stress » et « prominence »

Ces considérations sur « stress » et « accent » nous amènent tout naturellement à une discussion sur les différents sens qui ont été donnés à ces deux termes dans la littérature. Il est essentiel que nous passions par ces points de terminologie car on ne peut parler de

la notion d'emphase sans avoir recours aux termes « accent », « stress » et « prominence ».

Hayes (1994 : 70) écrit par exemple:

To emphasize a constituent, exaggerate (to the desired degree) the phonetic correlates of stress within that constituent.

Bolinger parle de « accent », pour d'autres encore le terme proéminence est parfois synonyme d'emphase.

Pour ce qui est du terme proéminence, un consensus semble tout de même exister. Une syllabe proéminente est une syllabe saillante par rapport aux autres, mise en relief par rapport aux autres :

Prominence is the property by which linguistic units are perceived as standing out from their environment (Terken, 1991 : 1768)¹⁰

Elle peut donc être emphatique ou pas. Seul Bolinger (1958 : 112) est en quelque sorte en marge car il définit la proéminence en terme d'intonation seulement :

a rapid and relatively wide departure from a smooth and undulating contour.

Mais là aussi selon cette définition, un élément proéminent peut être emphatique ou non.

Nous utiliserons le terme proéminence de la façon la plus traditionnelle. La notion de proéminence est une notion très large, qui peut englober n'importe quel type de mise en relief, emphatique ou non, intonationnelle ou non.

Pour les termes « stress » et « accent », des divergences importantes existent selon les auteurs et les termes ont évolué ces dernières années.

1.1.4.1 « Stress-as-loudness »

Dans une succession de syllabes ou de mots prononcés, certains apparaissent plus saillants, plus proéminents que d'autres. C'est à ce phénomène que l'on donne traditionnellement le nom de « stress ».

¹⁰ Cf. aussi Campbell, 1995a par exemple.

A quoi cette proéminence est-elle due ? Là est la réelle question qui a causé tant de dissensions parmi les phonéticiens et linguistes. Déjà Sweet (1906) employait le terme « stress » et l'associait à la force¹¹ :

The comparative force with which the syllables that make up a longer word are uttered is called 'stress'. (Sweet, 1906 : 49)

Les linguistes de l'époque associaient donc « stress » et force, ce qui servait de base à la taxinomie qui séparait des phénomènes prosodiques impliquant la force de ceux impliquant la hauteur mélodique. Passy (1906) parle d'accent de force et d'accent musical. Cette dichotomie a perduré, en passant par Bloomfield (1933 : 110-111) :

Stress – that is, intensity or loudness – consists in greater amplitude of soundwaves, and is produced by means of more energetic movements such as pumping more breath, bringing the vocal chords closer together for voicing and using the muscles more vigorously for oral articulation.

et Kingdon (1958a ;1958b) qui distingue « word stress » (dont l'équivalent français est « accent lexical »), c'est-à-dire le degré relatif de force donné aux différentes syllabes d'un mot polysyllabique, et « sentence stress » (ou encore « accent de phrase » en français), c'est-à-dire la force relative donnée aux différents mots d'une phrase. Ainsi un mot monosyllabique n'a pas d'accent lexical mais peut porter un accent de phrase. C'est à ces définitions que la littérature se réfère traditionnellement. Kingdon incorpore la notion de hauteur mélodique et distingue les tons statiques et les tons kinétiques. Les tons kinétiques sont les syllabes dont la proéminence est due à un mouvement mélodique, descendant (ˊ), montant (ˋ), montant-descendant (ˊˋ), ou descendant-montant (ˋˊ). Les tons statiques sont des syllabes qui sont proéminentes car placées sur un ton haut (ˊ) ou bas (ˋ) par rapport aux autres. A propos des tons statiques, Kingdon (1958b : 2) précise :

[F]rom the stress point of view, they are prominences, but [...] from the intonation point of view, they make no contribution to the meaning or feeling of an utterance

alors que les tons kinétiques sont « moving and meaningful » (1958a : xxv) et que leur fonction est d'exprimer des sentiments (1958b : 2). On aborde ici des notions intéressantes mais des études expérimentales ont montré que tout cela était faux. Les

¹¹ Par force, Sweet (1906 : 47) entend l'effort avec lequel l'air est expulsé des poumons. La force se traduit acoustiquement par ce que l'on appelle en anglais « loudness ».

syllabes accentuées n'ont pas forcément révélé un pic d'intensité plus fort que les syllabes inaccentuées. Il fallait donc revoir ces concepts qui semblaient pourtant faire l'unanimité.

1.1.4.2 « Stress-as-accent »

Au vu de ces expériences et d'autres, qu'il a lui-même réalisées, Bolinger (1958) argumente que « stress » est donc réalisé par des mouvements mélodiques et remplace le mot « stress » par « pitch accent » (ou « accent ») qu'il définit comme :

Prominence due to the configuration of pitches. (Bolinger, 1958 : 36)

Il pense que le seul type de proéminence possible est un accent mélodique. Trois sortes d'accents existent selon lui (cf. Bolinger, 1958) :

- un mouvement mélodique descendant après la syllabe proéminente (A) ;
- un mouvement mélodique montant vers ou après la syllabe proéminente (B) ;
- un mouvement mélodique descendant vers la syllabe proéminente (C).

Ce qu'il appelle « lexical stress » n'est pas selon lui une proéminence et devient une notion abstraite. Il s'agit du lieu potentiel pour un accent mélodique. Ainsi « pitch prominence » et « lexical stress » sont associés dans la réalisation de « accent », qui est sémantique : une phrase va changer de sens selon que l'on place l'accent à tel ou tel endroit.

Selon Crystal (1969), si l'intensité est le paramètre principal dans la perception d'une proéminence, celle-ci est alors appelée « stress ». Si c'est la fréquence fondamentale qui est le critère le plus fort, Crystal parle alors de « accent ». Il se rapproche des linguistes du début du siècle mais avec cette différence qu'il conçoit que l'on puisse avoir un mouvement mélodique dans les deux cas. C'est le critère du paramètre le plus important qui l'emporte. Mais il estime que l'on peut aussi percevoir une syllabe proéminente sans mouvement mélodique. C'est le cas dans le chuchotement par exemple. Alors c'est l'intensité qui donne l'impression de proéminence, mais Crystal parle aussi de durée, ce qui nous semble pertinent.

Ce qui était traditionnellement appelé « sentence stress » devient donc, dans la terminologie de Crystal, « accent » dans les parties où l'on trouve une proéminence

mélodique et « stress » dans les parties de l'énoncé où l'on n'a pas de proéminence mélodique. L'accent lexical (« lexical stress ») devient pour lui aussi une notion abstraite qui n'appartient pas au domaine de l'analyse prosodique mais au lexique.

Crystal distingue plusieurs degrés de proéminence dus à l'intensité : « strongly stressed » ('), « stressed » (') et « unstressed », qui peuvent avoir différents degrés de proéminence dus à la hauteur mélodique : « primary, secondary or tertiary accent ». « Stress » et « accent » peuvent donc non seulement se combiner mais ils dépendent en fait l'un de l'autre : un « secondary accent » par exemple ne peut avoir lieu que sur une « stressed syllable ».

Le problème que pose cette théorie est que l'on n'a aucun moyen de distinguer une proéminence nucléaire d'une proéminence non nucléaire puisqu'à la fois les « stressed syllables » et les « accented syllables » peuvent être nucléaires ou non nucléaires.

1.1.4.3 « Stress-as-rhythm »

Ladd (1980 : 25ff), à la suite de Householder (1957), Abercrombie (1976) et Liberman (1975), propose une alternative à cela en parlant de rythme : « stress-as-rhythm ». Dans le cadre de cette théorie, les syllabes que nous percevons « stressed » ne sont autres que celles qui correspondent aux « rhythmic beats », et ceux-ci se retrouvent à intervalles de temps assez réguliers dans le discours, ou du moins sont perçus ainsi d'après le principe d'isochronie (cf. O'Connor, 1965 ; Lehiste, 1977)¹².

Cette théorie a l'avantage d'expliquer plusieurs phénomènes tels que le glissement accentuel (« stress-shift »), ou encore l'accent silencieux (« silent stress ») (cf. Couper-Kuhlen, 1986 : 33ff, pour plus de détails). Ainsi, ce qui est appelé « stress » n'aurait pas de rapport avec le mouvement mélodique par exemple dans le cas des accents silencieux, ni avec l'accent lexical pour le glissement accentuel par exemple. Tout ne serait qu'une question de rythme.

Couper-Kuhlen (1986 : 35) avoue qu'aucune de ces théories n'est entièrement satisfaisante mais tend tout de même en faveur de cette dernière.

¹² Cf. Classe (1939) ou Shen & Peterson (1964) pour des arguments à l'encontre du principe d'isochronie.

1.1.4.4 « Stress and non-stress accent »

Nous avons intitulé cette sous-section d'après l'ouvrage de Beckman (1986) car elle y développe une théorie de « stress » et « accent » en lien avec « tone » qui se détache des autres et met peut-être tout le monde d'accord. Beckman a étudié l'anglais et le japonais plus particulièrement mais également d'autres langues, à tons ou non. Beckman (1986 : 1) définit « stress » et « accent » de la façon suivante :

ACCENT : a system of syntagmatic contrasts used to construct prosodic patterns which divide an utterance into a succession of shorter phrases.

STRESS : a phonologically delimitable type of accent in which the pitch shape of accentual pattern cannot be specified in the lexicon but rather is chosen for a specific utterance from an inventory of shapes provided by the intonation system.

Et selon Beckman, le ton (« tone ») est différent de « accent » en cela que « tone » est une notion paradigmatique alors que « accent » est une notion syntagmatique. Sur un mot monosyllabique, on peut avoir différents tons qui donneront différents sens au mot, alors qu'on ne peut avoir qu'un accent. Le ton, qu'elle nomme finalement « non-stress accent », a une réalisation uniquement mélodique, alors que « accent » dans la terminologie traditionnelle, qui devient « stress accent » pour elle, est réalisé par la mélodie, mais pas uniquement : l'intensité et la durée entrent également en ligne de compte.

Ainsi Beckman distingue les « stress accent languages », comme l'anglais ou le néerlandais, et les « non-stress accent languages », comme le japonais ou le chinois mandarin.

Cette idée de « stress accent » réalisé par plusieurs paramètres à la fois attire notre attention et nous nous demandons si l'emphase n'est pas une forme de « stress accent » dont les paramètres seraient simplement décuplés.

1.1.5 Conclusion

Après ce survol des différentes utilisations des termes topique, focus, contraste, « accent », « stress » et proéminence, nous pouvons affirmer qu'il n'existe aucun consensus dans ce domaine. Chacun utilise ces termes à sa façon ou selon la tradition de

l'école à laquelle il appartient, et nous ne souhaitons pas dans le cadre de ce travail nous lancer dans un débat sur le bien-fondé de telle ou telle définition ou utilisation. Nous voulions simplement essayer de mettre un peu plus de clarté dans cette pénombre sémantique et surtout ne pas passer à côté de certaines études réalisées sur l'emphase ou qui s'y rapportent pour une question de terminologie. Nous savons maintenant qu'il nous faut nous intéresser à ce qui a été dit sur la prééminence par exemple, même si ce n'est pas l'équivalent exact d'emphase, sur le contraste et sur « stress ».

Nous pouvons également affirmer au terme de cette première sous-section qu'il faudra consacrer une grande partie de notre travail à l'importance de la place de l'accent nucléaire, puisqu'il semble s'agir là du débat essentiel (cf. focus large et focus étroit) ainsi qu'à l'importance du contexte sémantique (cf. information nouvelle et information ancienne) et prosodique (accents prénucléaires, unité intonative dans son ensemble).

Nous ne souhaitons pas non plus donner tout de suite une définition précise de l'emphase, le but de notre étude étant précisément de tenter de définir l'emphase sur un plan phonologique et même peut-être plus large. Nous devons cependant en donner une définition générale, voire vague. Cette sous-section ne nous ayant pas permis de dégager une définition précise, et confirmant ce que nous disions en introduction, à savoir que l'emphase est une notion utilisée par tous mais définie par personne, nous proposons pour point de départ notre propre définition de l'emphase, à la fois intuitive et inspirée de toutes les lectures décrites ci-dessus. Nous rejoignons la définition rhétorique de Bussman *et al.* (1996), ainsi que l'étude de Selting (1994) : l'emphase est pour nous une façon anormale, inattendue, de mettre une partie d'un énoncé en relief. Un énoncé ou une partie d'un énoncé est emphatique selon nous s'il est non neutre, si le locuteur l'a prononcé d'une façon qui marque une implication de sa part plus importante qu'à la normale.

1.2 *Etudes expérimentales*

Nous souhaitons à présent passer en revue les travaux empiriques qui ont été réalisés en relation avec l'emphase.

Comme nous l'avons vu, emphase et émotion sont intimement liées et Günthner (1996 : 292) parle en effet d'attitudes affectives et de charge émotionnelle en lien avec l'emphase :

emphasis can be used to signal a great variety of affective stances (such as interest, relevance, importance, unexpectedness and high positive as well as high negative emotional load).

C'est pourquoi nous commencerons par un survol des études entreprises sur l'émotion.

Nous verrons ensuite quelles études ont été menées sur « accent », « stress », et prééminence, maintenant que nous savons plus précisément de quelle façon ces notions sont liées à l'emphase.

Nous nous pencherons pour finir sur les études qui ont trait plus précisément à l'emphase seule (et au contraste quand il est entendu comme l'équivalent exact d'emphase par les chercheurs).

Cette revue des études expérimentales liées à l'emphase nous permettra de mieux définir les domaines à étudier pour notre propre étude afin de contribuer à la recherche en essayant de ne pas réitérer les expériences déjà réalisées.

1.2.1 Emotion

Plusieurs expériences ont pour objet les liens entre intonation et émotion ou attitudes. On peut classer ces études en trois groupes.

1.2.1.1 Paramètres acoustiques et attitudes

Une première approche consiste à faire des investigations expérimentales pour relier des paramètres acoustiques et des attitudes.

C'est le choix de Lieberman & Michaels (1962), qui ont enregistré six locuteurs prononçant huit phrases différentes exprimant une émotion différente trois fois chacune : une phrase exprime l'ennui, une autre est dite de façon confidentielle, une question exprime le doute ou l'incrédulité, un message exprime la peur, une question est posée de façon objective, une phrase énoncée de façon objective, et la dernière est censée sonner « pompeuse ». A l'aide d'auditeurs, les meilleurs exemples des différentes émotions exprimées ont été sélectionnés. Ces échantillons ont été synthétisés et l'on a apporté les modifications nécessaires pour faire écouter à des auditeurs six types de phrases :

- 1) l'enregistrement original ;
- 2) le contour intonatif seulement sans l'amplitude ;
- 3) le contour intonatif avec les modulations correctes de l'amplitude ;
- 4) le contour intonatif avec l'amplitude mais avec 5 variations de la fréquence fondamentale lissée sur la base d'intervalles de 40 millisecondes ;
- 5) la même chose exactement mais avec des intervalles de 100 millisecondes ;
- 6) l'amplitude seulement avec une hauteur constante.

Les auditeurs devaient identifier les huit émotions à partir de ces enregistrements. Les résultats sont les suivants :

- ⇒ 1) : 85% d'identification correcte ;
- ⇒ 2), 3), 4) : 40-45% d'identification correcte ;
- ⇒ 5), 6) : 25 et 14 % d'identification correcte.

Lieberman & Michaels (1962 : 927) en ont conclu qu'on ne peut faire ressortir un seul corrélat acoustique mais qu'une combinaison de tous les paramètres étudiés est nécessaire pour faire passer une émotion :

There is no single acoustic correlate of the emotional modes of this experiment. Phonetic content, gross changes in fundamental frequency, the fine structure of the fundamental frequency, and the speech envelope amplitude, in that order, all contributed to the transmission of the emotional modes.

Ils notent aussi que le critère de durée ne fait pas partie de leur expérience. Et enfin ils expliquent que les auditeurs ne sont pas tous sensibles aux mêmes paramètres acoustiques en ce qui concerne la transmission des émotions.

1.2.1.2 Contours intonatifs et émotions ou attitudes

Une deuxième approche consiste à essayer de faire coïncider des émotions choisies et des contours intonatifs, de façon expérimentale également. C'est le cas des travaux de Uldall (1964) ou encore de Osser (1964).

Uldall (1964) a choisi de faire enregistrer cinq phrases à un locuteur (une phrase déclarative, une yes-no question, une question avec un pronom interrogatif, un ordre et une phrase vide de sens), puis ces phrases ont été synthétisées avec 16 contours intonatifs différents pour chacune. On a alors demandé à un groupe de 15 auditeurs de placer ces phrases sur des échelles d'adjectifs s'opposant (échelles empruntées à Osgood : cf. Osgood *et al.*, 1957) :

bored-----interested

polite-----rude

timid-----confident

etc...

En construisant ce qu'elle appelle des « diagrammes d'espace sémantique », Uldall postule qu'il existe trois dimensions de sens et en arrive aux conclusions suivantes :

- on peut donner une description de chaque contour intonatif ;
- certaines émotions sont clairement exprimées par l'intonation (toujours la même) ;
- plus le contour est « vivant », plus sa position dans l'espace sémantique va être stable :

the more 'lively' the contour is, the more stable is its position in the semantic space over the different sentences. (Uldall, 1964 : 279)

Uldall note enfin que le tempo ou la qualité de la voix, ainsi que les effets contextuels, qu'elle n'a pas pris en compte dans son expérience, seraient à étudier.

Osser (1964) a, quant à lui, enregistré une actrice prononçant le mot « home » (comme en réponse à la question « Où allez-vous ? ») de 10 manières différentes exprimant la colère, l'inquiétude, l'ennui, le calme, l'enthousiasme, la joie, l'amour, une question, la tristesse, et de façon la plus neutre possible. Il a ensuite décrit chaque énoncé à l'aide des phonèmes définis par Trager & Smith (1951) et également à l'aide d'un système de traits distinctifs paralinguistiques du type voix traînante, trop forte, grinçante, etc... On

a demandé à des auditeurs de relier chaque énoncé avec une des 10 catégories en autorisant des doublés. Osser a construit à partir de là des matrices de confusion pour les différentes émotions et a tenté de relier le degré de confusion et le degré de similarité (la similarité étant exprimée au moyen de contours intonatifs et de son système de traits distinctifs paralinguistiques). Ses conclusions l'ont amené à dire que son système de description paralinguistique prédisait bien la confusion chez les auditeurs alors que la description avec les phonèmes de Trager & Smith était nettement moins satisfaisante.

Ces deux approches (celle exemplifiée par Lieberman & Michaels, 1962, et celle choisie par Uldall, 1964, ou Osser, 1964) ont été critiquées, entre autres par Ladd (1980 : 131) qui déplore que l'on parte du sens pour aller à la forme et non l'inverse :

They work from meaning to form rather than the other way round. They underemphasize the contribution to the expression of attitude made by lexical and grammatical choices and by the congruence or incongruence of all the attitudinal cues,

ou encore Selting (1994 : 378) qui regrette que ces expériences ne fournissent que des catégories prosodiques et attitudinales préconçues et qu'elles ne permettent de tirer des conclusions qu'à propos de réactions conditionnées par les auteurs de l'étude sur des phrases créées pour l'étude :

They only allow statements about the subjects' (forced-choice) reactions to the researchers' suggested, preconceived, and most often introspectively based, hypotheses.

Elle ajoute que ces études ne permettent pas de tirer des conclusions sur la parole spontanée :

[these experiments] do not allow any statements about subjects' and speakers' uses or interpretations of prosodic categories in natural communicative situations. (Selting, 1994 : 378)

On peut aussi reprocher à ces études de ne pas prendre en compte la contribution des choix grammaticaux et lexicaux dans la façon de faire passer les émotions, comme Ladd le mentionne (citation ci-dessus : Ladd, 1980 :131).

O'Connor et Arnold (1973) ont essayé de déterminer empiriquement quelles attitudes (« attitudinal meanings ») chaque groupe intonatif (« tone group » ou « nuclear tone ») exprime. Couper-Kuhlen (1986 : 179) donne un résumé des attitudes liées aux différents groupes intonatifs, par exemple :

Tone 5 (basically a falling-rising tune with optional low pre-head and tail, for instance, is said to convey the following attitudes in conjunction with statements : 'grudgingly admitting, reluctantly or defensively dissenting, concerned, reproachful, hurt, reserved, tentatively suggesting ; (in echoes) greatly astonished'.

Cette approche a aussi fait l'objet de critiques car elle avait tendance à utiliser des phrases « trop bien choisies ». Crystal (1969 : 294ff) a déploré le nombre illimité d'attitudes décrites.

D'autres études font une distinction entre des intonations marquées et non marquées (Halliday, 1967a). Les intonations non marquées ont des sens plutôt grammaticaux alors que seules les intonations marquées sont reliées à l'expression de l'émotion (Crystal, 1969 : 261). Cela n'est pas pour plaire à Bolinger (1986 : 27) qui estime que la fonction principale de l'intonation est la transmission des émotions :

though intonation is indispensable to grammar, the grammatical functions of intonation are secondary to the emotional ones.

Selting (1994), qui a étudié elle aussi les usages marqués et non marqués des paramètres prosodiques sur les questions, indique que les usages marqués indiquent une implication plus grande de la part du locuteur :

the marked use of prosodic parameters, i.e. a use that is recognizable as more noticeable and salient than in the surrounding units, signals some sort of 'special' meaning. In the appropriate context, such special marked uses of prosody may be interpretable as signalling heightened involvement or emotive meaning. (p. 381)

Ceci nous rappelle bien entendu la définition que donne Bussman *et al.* (1996) de l'emphase¹³.

¹³ Cf. 1.1.1.

Nous reviendrons sur cette distinction plus loin en relation avec des études portant plus précisément sur l'emphase¹⁴.

1.2.1.3 Travaux récents

Encore plus récemment, un véritable engouement pour l'étude de l'émotion s'est fait jour, ceci dû à l'avancement des travaux sur la synthèse de la parole. En effet, les émotions sont étudiées d'un point de vue acoustique et phonologique afin de pouvoir trouver des paramètres susceptibles d'être intégrés dans un modèle de parole automatique.

C'est exactement le cadre de l'étude de Mozziconacci (1998) par exemple, qui a étudié les paramètres prosodiques suivants : hauteur mélodique, dynamique fréquentielle, configuration intonative et débit de parole pour sept émotions. Encore une fois, les sept émotions sont déterminées à l'avance et on demande à des locuteurs (ou acteurs) de lire des textes avec ces différentes émotions. Ceci n'est donc pas de la parole spontanée. L'étude de Mozziconacci se rapproche fortement de celle de Uldall (1964, mentionnée plus haut). Les conclusions que tire Mozziconacci sont cependant intéressantes :

une interaction de paramètres prosodiques permet l'expression orale de l'émotion et la plupart des émotions peuvent être véhiculées dans la parole de synthèse grâce aux paramètres ici étudiés. (1998 : 207)

Elle ajoute que les résultats sont moins convaincants pour certaines émotions et que :

il se peut que d'autres facteurs tels que la qualité de la voix, son intensité ou d'autres propriétés de l'intonation soient essentiels. (Mozziconacci, 1998 : 207)

Trouvain et Barry (2000) ont étudié la prosodie de l'excitation dans les commentaires de courses de chevaux. Les paramètres qu'ils ont choisi d'analyser sont les pauses, la respiration, la fréquence fondamentale, l'intensité et l'inclinaison du spectre et leur étude montre finalement que les commentaires de courses de chevaux et donc la prosodie de l'excitation sont caractérisés par des pauses plus fréquentes et plus courtes, une respiration accélérée, une fréquence fondamentale beaucoup plus haute, une plus

¹⁴ Cf. 1.2.3, Tannen (1984) par exemple.

forte intensité et un spectre plus plat (p.89), les caractéristiques restant tout de même très individuelles.

L'étude de Hirschberg & Ward (1992) sur le contour intonatif « rise-fall-rise » en anglais (L*+HLH%, selon la théorie de Pierrehumbert, 1980) ne porte pas exactement sur les émotions mais s'intéresse aux interprétations d'incrédulité et d'incertitude, que nous pourrions ramener à des émotions, c'est pourquoi nous avons choisi de mentionner cette étude ici. Hirschberg & Ward ont pris huit paires d'énoncés dont les phrases et les contours sont identiques mais le contexte différent afin d'induire les interprétations « incrédulité » et « incertitude » (d'après des recherches antérieures, cf. Ward & Hirschberg, 1985 ; 1988). La durée, l'amplitude, la hauteur mélodique, et les caractéristiques spectrales pour les deux interprétations ont été mesurées, montrant des différences dans les quatre domaines et également dans la qualité de la voix (Hirschberg & Ward, 1992 : 245). Afin de déterminer quelles distinctions phonétiques sont déterminantes dans la perception des auditeurs pour interpréter les phrases, les traits pour les phrases avec l'interprétation 'incrédulité' ont été substitués aux traits correspondants dans l'autre interprétation, de façon à produire des stimulus avec toutes les combinaisons possibles des quatre paramètres, ceci donnant seize stimulus pour chaque paire. Des auditeurs devaient ensuite dire quelle interprétation ils en donnaient à chaque fois. Les résultats de cette étude montrent que le facteur le plus déterminant dans la perception des différentes interprétations est la variation de la hauteur mélodique, c'est-à-dire la fréquence fondamentale, suivie de loin par les caractéristiques spectrales. La durée et l'amplitude s'avèrent n'avoir aucun effet, malgré les différences trouvées pour les deux interprétations. A l'instar de Bolinger (1986), Hirschberg & Ward (1992 : 250) parlent alors d'implication de la part de l'énonciateur :

[L]arger pitch ranges indicate a greater degree of involvement and smaller pitch ranges indicate a smaller degree. Intuitively, incredulity would appear to entail a greater involvement on the part of the speaker than would uncertainty.

et encouragent des recherches futures dans ce sens sur d'autres contours intonatifs. Nous notons qu'il est question ici à nouveau d'implication de la part du locuteur.

Si l'on veut aller encore plus loin en ce qui concerne l'étude des émotions, on peut citer des travaux sur la gestuelle, avec Massaro (1987) et ses têtes parlantes ou encore Tarter & Braun (1994) dans une étude sur le sourire ou bien Guaïtella (1995) qui parle de la mélodie du geste, et a beaucoup étudié les mouvements des sourcils (cf. aussi Cavé *et al.*, 1996, et Purson *et al.*, 1999).

1.2.2 « Stress » et proéminence

Nous souhaitons examiner dans cette section non pas tous les travaux qui ont été faits sur « stress » (qui pourrait être traduit en français par « accent lexical » dans la plupart des cas) ou sur la proéminence car cela s'éloignerait trop de notre propos et nous risquerions des redondances inutiles, mais uniquement les travaux les plus incontournables et ceux qui nous semblent novateurs et mettent en valeur des paramètres différents de ceux habituellement analysés, ces paramètres pouvant être utilisés pour la détection de l'emphase. Nous emploierons ici le terme français « accent » pour faire référence à l'anglais « stress », dans le sens de accent lexical. Nous dirons « accent mélodique » pour ce que l'on nomme traditionnellement en anglais « accent » ou « pitch accent ».

1.2.2.1 Durée, intensité et fréquence fondamentale

Fry (1955 ; 1958) a beaucoup travaillé sur ce qu'il appelle « linguistic stress ». Il s'agit de ce que l'on appelle en français l'accent lexical : dans son étude de 1955, il a choisi des paires de dissyllabes nom/verbe pour lesquelles l'accent passe de la première syllabe quand c'est le nom à la deuxième syllabe quand il s'agit du verbe. Il a analysé la durée et l'intensité et a effectué des tests de perception en faisant varier ces paramètres. A la fois la durée et l'intensité se sont révélées des paramètres essentiels dans la perception de l'accent, la durée étant un paramètre un peu plus effectif. Dans son étude de 1958, Fry reprend les mêmes paires de dissyllabes et analyse cette fois les quatre paramètres acoustiques suivants : durée, intensité, fréquence fondamentale et structure de formants, car ce sont les corrélats physiques des facteurs perceptifs que sont la

longueur, la force, la mélodie et la qualité. Plusieurs tests ont été réalisés en synthétisant les stimulus et en faisant varier les différents paramètres. Les résultats confirment ceux de 1955, c'est-à-dire que la durée et l'intensité restent des paramètres essentiels dans la perception de l'accent, surtout la durée, mais permettent d'affirmer que la fréquence fondamentale est un paramètre encore plus important.

A la suite de Fry, des linguistes ont fait des recherches similaires et Terken (1991), qui a étudié très précisément les relations entre la fréquence fondamentale et la perception de la proéminence, ou encore Brown & McGlone (1974), qui ont mesuré la pression de l'air intraoral, le flux d'air, l'intensité vocale, la fréquence fondamentale et la durée des phonèmes accentués, confirment que la fréquence fondamentale est le paramètre le plus important.

Weismer & Ingrisano (1979) ont étudié le facteur durée et arrivent eux aussi à la conclusion que les mots mis en relief subissent un allongement et que lorsque la mise en relief apparaît vers le début de la phrase, il y alors un raccourcissement des éléments qui suivent.

Bagshaw (1993 : 333) décrit

un algorithme qui transforme des paramètres acoustiques du signal sonore en éléments abstraits permettant de transcrire automatiquement l'accent de phrase et les mouvements intonatifs.

Il utilise la durée, l'énergie et la fréquence fondamentale et prend comme unité de référence la syllabe. La proéminence de chaque syllabe est déterminée de façon automatique et la transcription résultante est comparée à une transcription manuelle. Le taux de concordance n'étant que de 61,6%, Bagshaw conclue que l'étiqueteur humain utilise probablement d'autres paramètres que ceux pris en compte par l'algorithme.

Selon Beckman & Edwards (1994) la proéminence est un système unidimensionnel avec trois niveaux qualitatifs :

- l'accent le plus fort est sur la syllabe qui contient une voyelle pleine et un accent mélodique nucléaire :

the last and strongest prominence in the constituent 'intermediate phrase';

- en deuxième position la syllabe contient une voyelle pleine mais pas de mouvement mélodique :

the first and strongest prominence in the constituent 'stress foot';

- et au dernier niveau, c'est-à-dire inaccentué, les voyelles sont réduites :

a non head element in the stress foot.

Beckman & Edwards mentionnent un quatrième type de proéminence (qui viendrait juste après le premier), une voyelle pleine et un mouvement mélodique prénucléaire (« prenuclear accented syllable »), mais ne l'incluent pas dans leur modèle.

Grâce à ce système, Beckman & Edwards expliquent pourquoi différentes études sur l'accent arrivent souvent à des résultats contradictoires (certains mettent en avant la F0 alors que d'autres parlent de durée¹⁵, d'intensité, ou encore d'autres paramètres acoustiques, ou mettent en lumière l'importance des consonnes¹⁶) et réconcilient tous les linguistes : la fréquence fondamentale est le paramètre le plus important quand il s'agit d'un accent de premier niveau et le devient de moins en moins jusqu'à ne plus l'être du tout tout en bas de l'échelle. Elle est alors remplacée par tous les autres paramètres tels que la durée, l'intensité, la qualité de la voyelle... :

At the lower level, the phonological content of the contrast is one of vowel quality. [...] The duration difference too is an inherent part of the specification of the prominence contrast at this level. At the higher level on the other hand, any effect of the prominence on the syllable's duration or on its degree of vocal-tract opening is ancillary to the associated intonation pattern. (p.30)

1.2.2.2 Déclinaison de la F0 et accent prénucléaire

L'effet de déclinaison de la fréquence fondamentale, c'est-à-dire l'abaissement progressif de la fréquence fondamentale tout au long d'un groupe intonatif, a été observé dans plusieurs langues (cf. Liberman, 1975 ; Maeda, 1976 ; Cooper & Sorensen,

¹⁵ Harrington (<<http://www.ling.mq.edu.au>>) met en avant le rôle de la durée des voyelles mais aussi des consonnes : « speakers vary the timing of their consonants and vowels in subtle ways to communicate word salience ».

¹⁶ Pierrehumbert & Talkin (1992) parlent d'un renforcement des traits consonantiques sur les syllabes accentuées à tous les niveaux.

1981 pour l'anglais¹⁷). Pierrehumbert (1979) s'est demandé comment ce phénomène était perçu par les auditeurs. Elle a présenté à des auditeurs des phrases contenant deux accents mélodiques (deux pics de fréquence fondamentale) et leur a demandé de juger quel était l'accent le plus haut. Lorsque les auditeurs jugeaient que les deux accents avaient la même hauteur, il se trouve qu'en fait le second accent était plus bas que le premier. En d'autres termes, cela signifie que si les pics avaient eu une hauteur égale, le deuxième aurait été jugé plus haut que le premier. Pierrehumbert explique ce phénomène par le fait que l'auditeur s'attend en fait à une baisse de la fréquence fondamentale, ce qui est la norme, et son jugement est donc basé sur cette attente.

Gussenhoven & Rietveld (1988) ont réalisé une étude dans laquelle les phrases choisies comportaient deux pics accentuels, des phrases du type da-DAH-da-da-da-DAH-da. Ils ont trouvé de façon tout à fait prévisible une corrélation entre le niveau mélodique du deuxième accent et la perception de sa proéminence. Mais ils ont également trouvé que le niveau mélodique du premier accent affecte de façon tout à fait inattendue cette fois la perception de la proéminence du second. Si le niveau mélodique du premier accent est réduit, le deuxième accent est perçu comme moins proéminent aussi. Ceci tend à prouver que la proéminence est une fonction de la variation mélodique de la phrase tout entière et que la fréquence fondamentale n'est pas une variable indépendante qui signale la proéminence de chaque accent individuel.

Ladd *et al.* (1994) ont poursuivi cette étude et sont parvenus à des résultats encore plus complexes. Ils ont trouvé que l'effet décrit ci-dessus par Gussenhoven et Rietveld n'a lieu que lorsque le deuxième accent n'a pas une hauteur mélodique très élevée. Lorsque le second accent excède un certain niveau (autour de 145 Hz pour la voix d'homme qu'ils ont utilisée), l'effet est inverse et une baisse de la hauteur mélodique du premier accent induit une perception plus grande de la proéminence du deuxième accent. Ladd *et al.* donnent alors une explication à ce phénomène : quand les accents sont normaux (c'est-à-dire non emphatiques), les variations mélodiques sont prises en compte pour la phrase tout entière et l'effet décrit par Gussenhoven et Rietveld se produit. Lorsque les accents sont emphatiques, les variations mélodiques sont évaluées accent par accent et

¹⁷ Cf. entre autres Vaissière (1983) pour le français, Thorsen (1980) pour le danois, Fujisaki & Nagashima

le fait de réduire la hauteur mélodique d'un accent en revient à augmenter celle de l'accent adjacent.

On note ici une fois encore le lien entre proéminence et emphase. Nous verrons plus loin¹⁸ qu'il est reconnu par d'autres chercheurs que la façon d'amener l'emphase - c'est-à-dire la façon de prononcer les mots et syllabes qui précèdent le pic emphatique - est importante dans sa perception, et que même ce qui suit le pic emphatique est parfois révélateur aussi.

1.2.2.3 Perception et tilt spectral

Campbell (1995a) a travaillé sur la proéminence, qui peut être une des définitions de l'emphase comme nous l'avons vu, et plus précisément sur la perception de la proéminence. Les paramètres que Campbell a choisi d'analyser sont novateurs. Il s'est penché sur les informations que donne le spectre (cf. aussi Campbell & Beckman, 1995) et affirme que

for dialogue speech, spectral information can be very helpful in discriminating prominences. (Campbell, 1995a : 678)

C'est l'avis de Sluitjer & van Heuven (1993 ; 1996), qui sont parmi les premiers à s'être intéressés à l'information spectrale en ce qui concerne la perception de la proéminence. Ils ont travaillé sur le néerlandais mais leur théorie est valable pour l'anglais aussi, les deux langues étant des « stress-accent languages » (par opposition à d'autres langues comme le japonais par exemple, qui sont des langues à « pitch accent » : cf. Beckman, 1986 et 1.1.4.4). Ils partent du principe connu que l'effort vocal est plus important pour réaliser une syllabe accentuée et que donc les fréquences hautes augmentent plus que les fréquences basses :

If a human speaker expends more vocal effort, as is required for the realization of stress, intensity does not change uniformly across the spectrum, but higher frequencies are increased more than lower frequencies. (Sluitjer & van Heuven 1993 : 246)

(1969) pour le japonais.

¹⁸ Cf. 1.2.3., Selting, 1994 ; Uhmman, 1992.

L'analyse de Sluitjer & van Heuven (1993) consiste alors à étudier les effets de l'accent lexical sur l'intensité. Dix locuteurs ont produit une phrase dans laquelle se trouve une paire minimale pour l'accent (un dissyllabe accentué tantôt sur la première syllabe, tantôt sur la deuxième et dont le sens diffère) en néerlandais. Sluitjer et van Heuven ont également remplacé ce mot par un mot réitérant (la même syllabe est répétée, ici « nana ») ne voulant rien dire. Les mots sont insérés dans une phrase normale et prononcés avec et sans mouvement mélodique sur la syllabe accentuée (« in and outside focus »).

L'intensité a été mesurée dans quatre bandes de fréquence contiguës. Les résultats montrent que les différences d'intensité entre les voyelles accentuées (« stressed ») et inaccentuées (« unstressed ») sont situées principalement au-dessus de 0,5 kHz, c'est-à-dire qu'un changement a lieu au niveau de la balance spectrale, qui met en relief les fréquences hautes des voyelles accentuées (la différence d'intensité va de 5 à 10 dB). Ce phénomène est encore plus fort si les voyelles accentuées se trouvent dans la portée du focus (« in focus » ou « accented » selon les termes des auteurs).

Il faut noter que les locuteurs ont été enregistrés dans des cabines de laboratoire avec la tête fixée à la chaise par une bande afin que la distance bouche – micro reste constante. Cette remarque est essentielle car en effet les variations d'intensité sont extrêmement sujettes aux bruits extérieurs, à la position du locuteur par rapport au micro, à ses mouvements, etc... On peut de ce fait se demander comment il est possible d'effectuer des mesures d'intensité en parole spontanée. En effet dans la conversation de tous les jours, des bruits interviennent sans arrêt, les locuteurs bougent, parlent en même temps ou pendant que l'un ou l'autre rit ou tousse par exemple. Nous reviendrons sur ce point (cf. 3.3).

La durée, l'intensité, et la qualité des voyelles (les formants) ont été également mesurées et la durée s'est révélée être le plus fort corrélat acoustique de l'accent lexical, alors que l'intensité et la qualité des voyelles ne sont que de minces indices selon Sluitjer et van Heuven. La balance spectrale est un corrélat presque aussi fort que la durée.

Il semble donc que les informations données par le spectre et en particulier ce qu'on appelle « spectral tilt » (balance ou inclinaison du spectre) soient en effet essentielles

dans la perception de la proéminence et donc également de l'emphase. Sluitjer & van Heuven (1993, 1995) sur le néerlandais et Campbell (1995a) et Campbell & Beckman (1995) sur l'anglais l'ont démontré dans leurs études. Le haut du spectre est renforcé quand il y a emphase sur une syllabe. C'est au niveau de la perception que cela a un rôle très important. L'intensité est accrue dans certaines fréquences que l'oreille perçoit et cela donne l'impression d'emphase, de mise en relief.

1.2.3 Emphase

Penchons-nous à présent sur les études plus précisément ciblées sur l'emphase ou le style emphatique ou encore « high involvement style ». Nous souhaitons commencer par rendre hommage à Coleman (1914).

1.2.3.1 Coleman

Il s'agit en effet de la plus ancienne référence que nous avons trouvée et il est absolument stupéfiant de voir à quel point Coleman était avancé dans ses recherches et de voir qu'il était surtout doté d'une oreille et d'une intuition exceptionnelles. En effet depuis 1914, les recherches sur l'emphase n'ont finalement pas fait beaucoup plus que tenter de trouver des justifications empiriques et scientifiques aux dires de Coleman. La machine a remplacé l'oreille pour plus de précision et surtout pour confirmer les hypothèses de l'époque.

Coleman donne une définition précise du terme emphase. Il distingue deux sortes d'emphase :

celle qu'il nomme proéminence et qui sert à mettre un mot en contraste avec un autre mot dans un autre groupe de mots, mais qui ne saurait en aucun cas se limiter à cette fonction contrastive. Il s'agit de ce que l'on a appelé dans la littérature l'emphase de contraste ou contrastive et que Coleman relie à l'apport d'information :

In nearly every sentence some word is regarded as the chief word (the *prominent* word) namely, that word which constitutes the logical predicate – the informing word, the word denoting the fact to which one's attention is

called, the word that answers the actual or imagined question, the last word that one would sacrifice to save a halfpenny on a telegram. (p. 9)

et celle qu'il appelle « intensity » et qui ne contraste pas un mot avec un autre mais qui sert simplement à faire une mise en relief. Il s'agit de l'emphase connue dans la littérature comme l'emphase d'intensité :

that manner of utterance which imparts an added degree of intensity to some part of the idea represented by a word. (p. 11)

Les deux types d'emphase peuvent être cumulés :

either kind of emphasis may be added to the other. (p.17)

Coleman fait ensuite une étude intonative, à l'oreille, de ces deux types d'emphase et il distingue « stress » (accent lexical) et « pitch » (hauteur mélodique). Mais surtout, il note déjà que l'intonation n'est pas le seul moyen d'exprimer l'emphase et que d'autres paramètres acoustiques, sémantiques, syntaxiques peuvent être ajoutés :

Intonation cannot, however, be regarded as the essential of intensifying expression. There enter into it a number of other factors – special stress, extra slowness, extra quickness, length of word, additional words before the intensified word to gain attention by keeping one waiting, pauses with the same object, and other devices, such as repetition or additional words generally. (p. 15)

C'est finalement dans cette direction que toutes les recherches sur l'emphase se sont tournées, et continuent de le faire, comme nous allons le voir.

Jones (1918) a suivi les traces de Coleman, il le reprend d'ailleurs beaucoup et il est intéressant de citer un passage dans lequel il associe un certain type d'emphase à certains mots, ce qui montre qu'il pense que l'emphase s'étend sur le domaine de la sémantique :

Contrast emphasis may be applied to almost any word, but intensity emphasis can only be applied to certain words expressing qualities which are measurable, e.g. adjectives such as huge, enormous, lovely [...], adverbs such as particularly, extremely, hopelessly, plural nouns such as [...] heaps, tons, hundreds, and a certain number of verbs such as rush, squeeze, hate. (p. 298)

1.2.3.2 Paramètres acoustiques

Wells (1986) a étudié les paramètres acoustiques liés à la perception de « focus » en anglais. Une première expérience de perception lui permet d'affirmer que l'on peut distinguer quatre catégories de « focus » en anglais :

- focus contrastif (« contrastive focus ») qui est alors unique et correspond au degré de prééminence maximal ; cette catégorie de focus a souvent été donnée équivalente à de l'emphase (cf. 1.1.2.) ;
- focus principal (« main focus ») qui est le focus primaire et correspond à un degré majeur de prééminence ;
- focus subsidiaire (« subsidiary focus »), secondaire ou tertiaire avec un degré mineur de prééminence ;
- focus zéro (« zero focus »), c'est-à-dire pas de focus et un degré de prééminence minimal.

Il analyse ensuite les paramètres acoustiques suivants :

- ⇒ (1) hauteur mélodique du point le plus haut du contour intonatif de la phrase (« pitch peak ») ;
- ⇒ (2) contour et hauteur mélodiques de tout le constituant mis en relief (« maximum pitch range ») ;
- ⇒ (3) chute ou montée intonative se trouvant juste avant le pic intonatif et faisant qu'il y a une interruption dans le contour (« pitch obstruction (a) ») ;
- ⇒ (4) idem mais juste après le pic intonatif (« pitch obstruction (b) ») ;
- ⇒ (5) le ton kinétique (« kinetic tone ») ;
- ⇒ (6) le pic d'intensité (« loudness peak ») ;
- ⇒ (7) le crescendo (la montée) dans l'intensité à partir du constituant précédant le constituant mis en relief jusqu'au constituant mis en relief (« crescendo ») ;
- ⇒ (8) le decrescendo (la chute) dans l'intensité depuis le constituant mis en relief jusqu'au constituant suivant (« decrescendo ») ;
- ⇒ (9) les pauses avant et après l'élément focalisé (« preceding & following pause ») ;
- ⇒ (10) l'allongement anormal d'une ou des syllabe(s) du focus, de ce qui se trouve avant le focus et de ce qui se trouve après le focus (« drawl ») ;

⇒ (11) l'accélération du tempo avant le focus, après le focus et sur le focus (« allegro »).

Les résultats de son expérience et de ses analyses montrent que les paramètres (1), (2), (5), (6), (8), (9), et (10) ou (11) sont les paramètres pertinents dans la perception du focus contrastif surtout. Pour les autres catégories de focus (qui sont d'intérêt moindre dans notre étude), moins de paramètres entrent en jeu. Cette étude attire tout particulièrement notre attention et insiste à nouveau sur le fait que plusieurs paramètres sont associés.

O'Shaughnessy (1979) a étudié la fréquence fondamentale dans plusieurs types d'énoncés dont les énoncés emphatiques. Il a travaillé sur des énoncés isolés lus en anglais américain, et ses résultats montrent que les syllabes mises en relief ont une fréquence fondamentale plus élevée, alors que les syllabes adjacentes, et en particulier celles qui se trouvent après la syllabe emphatique montrent un abaissement de la fréquence fondamentale :

if there is any F0 feature that could be called 'emphasis', it would likely comprise an F0 contour with one or two large rise + fall obstrusions, each of about a syllable's duration, with other (especially ensuing) F0 relegated to a low level (O'Shaughnessy, 1972 : 124).

Forts des expériences précédentes où la fréquence fondamentale et la durée sont reconnues comme les paramètres les plus importants dans les éléments mis en relief, Cooper, Eady & Mueller (1985) ont souhaité aller un peu plus loin dans l'analyse de ces deux paramètres dans les énoncés contenant des éléments avec un accent contrastif (« contrastive stress ») en étudiant également les éléments adjacents aux éléments mis en relief. Leur première expérience porte sur des phrases enregistrées par des locuteurs américains en réponse à des questions qui font varier l'élément mis en relief. La fréquence fondamentale et la durée sont analysées et les résultats éloquentes.

La durée augmente sur les versions emphatiques, de façon localisée :

A word is significantly longer when it is emphasized than when it is not, and the presence of an emphasized item has no effect on the durations of other content words in a sentence. (Cooper *et al.*, 1985 : 2146)

De plus, l'augmentation de la durée sur un mot mis en relief dépend de la durée caractéristique de ce mot : plus le mot est long, moins la durée en sera allongée. Ceci n'est plus vrai lorsque le mot est en position finale car alors les effets de l'allongement dû à la position finale entrent en jeu.

Quant à la fréquence fondamentale, elle joue également un rôle déterminant, augmentant de façon significative sur les éléments mis en relief. Mais elle varie aussi sur les mots adjacents : ceux qui sont suivis d'un élément emphatique ont une F0 plus élevée que ceux qui sont précédés par un élément emphatique. On pourrait parler soit d'effet anticipatoire, soit d'effet rétroactif. Ces résultats corroborent donc ceux de O'Shaughnessy.

La deuxième expérience est un prolongement de la première, avec des phrases plus longues et des phrases ne contenant aucun élément mis en relief.

En ce qui concerne la durée, les résultats confirment exactement ceux de la première expérience et permettent d'affirmer que les mots non emphatiques qui se trouvent dans des phrases contenant un élément emphatique ont la même durée que les mots qui se trouvent dans des phrases neutres :

unemphasized words in sentences containing contrastive stress obtain duration values equal to those observed in neutral sentences. (*ibid.* : 2150)

Pour ce qui est de la fréquence fondamentale, les résultats corroborent également ceux de la première expérience, mais permettent de choisir entre l'hypothèse de l'effet anticipatoire ou celle de l'effet rétroactif. C'est ce dernier qui l'emporte car pour les phrases plus longues, le contour intonatif ne joue pas un rôle déterminant pour marquer un mot avec un accent contrastif en position finale. Ceci réfute donc l'hypothèse de l'effet anticipatoire :

This finding refutes the hypothesis that the F0 values on words preceding contrastive stress are increased in anticipation of the emphasized item. (*ibid.* : 2153)

Les chercheurs concluent finalement que le problème du focus est plus compliqué que l'on ne le croit et surtout que la durée et la fréquence fondamentale des éléments mis en relief varient en fonction de leur place dans la phrase. En positions initiale et médiane, l'effet le plus significatif est un abaissement de la F0 sur les mots suivant l'élément mis en relief, plutôt qu'une F0 plus haute sur l'élément emphatique :

for sentence-initial and sentence-medial focus, the major effect on F0 is a post-focus lowering rather than a heightening of F0 on the focused word itself. (*ibid.* : 2154)

Quant à la déclinaison de la F0 qui était aussi une hypothèse de départ, elle s'avère n'apparaître que dans les phrases lues de façon isolée sans contexte et est beaucoup moins évidente dans le discours spontané.

Eady, Cooper, Klouda, Mueller & Lotts (1986) ont continué leurs recherches et ont voulu étudier les différences entre focus étroit et focus large. Il existe en effet des divergences selon les linguistes (cf. 1.1.2.2.), mais aucune étude réellement empirique n'avait été réalisée.

Leur première expérience consiste à analyser les données acoustiques (durée et fréquence fondamentale à nouveau) de phrases ne différant que sur la portée du focus, focus neutre, large et étroit. Eady *et al.* ont travaillé sur des phrases lues, en réponse à différentes questions permettant de faire varier le focus (même principe que pour l'étude de 1985 décrite ci-dessus). Les analyses portent aussi sur les différences en fonction de la place de l'élément focalisé.

Les résultats montrent une nette différence entre focus étroit et focus large : le focus étroit est caractérisé par une F0 plus haute et une augmentation de la durée toutes deux localisées sur l'élément focalisé. Dans le cas du focus large, on observe une augmentation de la durée beaucoup plus étendue (sur plusieurs mots) et pas d'augmentation de la hauteur de la fréquence fondamentale.

La deuxième expérience teste des phrases à deux focus étroits. Là encore, les résultats pour ce qui est de la durée confirment qu'un élément mis en relief a une durée plus importante et que celle-ci varie en fonction de la place de l'élément dans la phrase (selon qu'il se trouve en position initiale ou en position finale, avec une longueur plus importante dans cette dernière position). De plus, on ne note pas une durée plus étendue sur les phrases à deux focus que sur celles à un focus. En revanche, et ceci est nouveau par rapport à l'étude précédente, Eady *et al.* notent un raccourcissement de la durée des éléments qui se trouvent après l'élément emphatique si celui-ci est en début de phrase. Ce n'est pas le cas si l'élément est en fin de phrase, évidemment.

Pour la fréquence fondamentale, les mêmes résultats que dans l'étude précédente sont avancés avec cette différence que lorsqu'il y a deux focus, on n'a pas cette chute de F0

après le focus que l'on observe dans les phrases à un seul focus en position initiale, certainement à cause de la présence du second élément focalisé en fin de phrase.

Ces expériences confirment encore qu'il ne faut pas s'en tenir à l'analyse des paramètres qui ne concernent que l'élément mis en relief mais que c'est l'unité intonative entière qu'il faut regarder et prendre en compte.

Tannen (1984) décrit ce qu'elle appelle « high-involvement style » (un style où l'implication est importante) en se basant sur du discours entre amis, ce qui est un grand pas en avant car la plupart des études précédentes se basaient sur des phrases prononcées hors contexte et choisies arbitrairement par l'auteur de l'étude. Ce qu'elle nomme une prosodie « expressive » est utilisée en co-occurrence avec d'autres stratégies selon elle. Parmi les paramètres qu'elle donne pour cette prosodie expressive, on trouve des formes marquées de « high and low pitches », l'utilisation de contrastes exprimés au moyen de changements marqués de hauteur et d'amplitude, une qualité de voix et des pauses marquées révélatrices. Tannen présuppose une forme non marquée, normale, qui n'est pas interprétée comme expressive ou marquant « high-involvement style ».

Quelques travaux récents et en particulier l'étude de Uhmman (1992) suggèrent que la densité de l'accentuation et la longueur des syllabes accentuées sont des moyens de marquer l'emphase.

Uhmman part de deux phénomènes :

- Density I : le nombre de syllabes par unité de temps.
- Density II : le nombre de syllabes accentuées par unité de temps.

Uhmman réalise elle aussi son expérience à partir de conversations de tous les jours et les résultats auxquels elle parvient sont les suivants :

⇒ une forte Densité I et une faible Densité II servent dans le cas de parenthèses, de « sidesequences », « afterthoughts », et de partie du discours de faible pertinence dans le développement de l'argumentation du locuteur. Ces passages sont perçus comme étant rapides ;

⇒ une faible Densité I et une forte Densité II sont réservées aux parties du discours qui ont une grande importance ou pertinence et sont perçus comme étant emphatiques et lents (cf. Uhmman, 1992 : 330).

Erickson et Lehisté (1995) ont étudié les durées des syllabes dans l'emphase contrastive. Leur étude consiste à faire prononcer un énoncé contenant des chiffres à une personne et l'interlocuteur feint de ne pas comprendre et amène donc la personne à répéter le bon chiffre. Cela se reproduit plusieurs fois. Erickson et Lehisté ont étudié la durée des syllabes sur lesquelles portaient l'emphase contrastive mais également la durée des autres syllabes. En comparant les durées trouvées avec les durées mesurées sur des phrases non emphatiques, ils montrent que l'emphase implique un allongement de la syllabe mise en relief mais également un raccourcissement des syllabes qui l'entourent. Ainsi l'emphase provoque une réorganisation temporelle de tous les mots d'une phrase. Il s'agit cependant d'un cas particulier d'emphase provoquée par la répétition et portant sur un chiffre. Or on sait que la prosodie des chiffres peut parfois constituer un cas bien précis¹⁹.

Selting (1994) a étudié le style emphatique (« emphatic speech style ») en allemand dans les narrations d'histoires où l'on arrive au point culminant. C'est-à-dire qu'elle considère que le point culminant est emphatique. Des locuteurs racontant des histoires ont été enregistrés en laboratoire. Selting analyse les paramètres prosodiques suivants :

- la densité de l'accentuation et le rythme (« dense accentuation and marked rhythm ») ; pour ce faire, elle compare le nombre et la longueur des syllabes accentuées par rapport aux syllabes inaccentuées dans les passages emphatiques et dans les passages non emphatiques ;
- les variations de hauteur mélodique et d'intensité (« global pitch and loudness variations ») ; elle mesure les hauteurs mélodiques des passages emphatiques et les compare à celles des passages non emphatiques et fait de même pour l'intensité ;

¹⁹ Cf. par exemple la discussion sur la prosodie des numéros de téléphone sur la liste de diffusion SproSIG (ISCA Special Interest Group on Speech Prosody, cf. références bibliographiques) du mois de janvier 2001.

- les variables liées aux accents localement marqués (« locally marked accent variants ») ; les pics de hauteur mélodique sont mesurés sur une syllabe bien précise, ainsi que l'intensité et l'allongement de la voyelle.

Au fil de son étude, Selting fait des remarques intéressantes et note qu'un des locuteurs combine des procédés prosodiques, lexicaux et syntaxiques qui constituent un style emphatique très marqué (p. 397). Elle souligne aussi l'importance des répétitions : « the repetition of the word 'todschick' signals emphasis », et des pauses qui rajoutent un effet dramatique (p. 398).

Elle conclut que les indices constitutifs du style emphatique sont les suivants :

- (1) forte densité de syllabes accentuées ;
- (2) une organisation rythmique « with short isochronous cadences » ;
- (3) des accents mélodiques marqués par un pic plus haut ou une intensité plus forte ;
- (4) des procédés lexicaux tels que l'intensification d'éléments lexicaux ;
- (5) des procédés syntaxiques tels que des ellipses ou des parallélismes.

On peut trouver 1) et 2) uniquement alors que les autres points sont forcément associés à 1) et 2), ce qui fait que Selting donne 1) et 2) comme les traits principaux du style emphatique :

in some cases, prosody is the only emphatic cue that makes climaxes in story telling recognizable and interpretable. (p. 405)

Cette étude de Selting est celle qui nous a semblé la plus complète mais nous lui reprochons de ne pas préciser sur quels critères elle détermine qu'un passage est emphatique ou non.

Winkler (1984) est parti d'une hypothèse tout à fait différente pour l'étude de segments emphatiques. Au lieu d'étudier les paramètres acoustiques classiques donnés par la plupart des linguistes quant à l'emphase (en particulier F0), il a décidé d'étudier les combinaisons de paramètres. Il part du principe que les variations de la fréquence fondamentale ne suffisent pas à expliquer pourquoi certains segments sont perçus comme emphatiques. Il postule l'existence d'un marqueur paraphonétique,

a sound which is not specific for language units like phonemes or intonemes.
(Winkler, 1984 : 327)

Ce marqueur fait qu'une partie de la phrase sonne « inhabituelle » (« unusual ») :

it manifests a specific character of the utterance which will be classified by the listener as 'emphatic' or 'conspicuous'. (*ibid.*)

Et comme le dit Lieberman (1974 : 2434), dans de tels cas, les auditeurs sont incapables de différencier les contours à un niveau linguistique alors qu'ils sont capables de leur donner des dimensions émotionnelles différentes :

In such cases, listeners are unable to differentiate the contours at any linguistic level though they may ascribe different emotional contexts to the contours. (*ibid.* : 329)

Selon Winkler, l'information paraphonétique est marquée par un changement dynamique de la combinaison phonématiquement nécessaire de paramètres acoustiques. Les paramètres acoustiques analysés sont la F0, l'intensité, le débit, les passages par zéro. Ainsi il note que certains phénomènes atypiques prennent place, comme la présence d'éléments voisés dans des consonnes sourdes par exemple :

emphatic speech simplifies, compensates and sometimes removes the specializations typical of phonemes (*ibid.* : 335).

Winkler conclut que l'emphase est rendue par une combinaison atypique de paramètres :

in emphatic sequences the combinations will be used freely, the neutral structures will be infringed, the concentration of curve types and parameters diminished and the special links between curve type/parameter or parameter/curve type changed or removed. (*ibid.* : 336)

Il rejoint la définition émise au départ de l'emphase :

'Emphatic' is the category for all sequences which seem to be non-neutral, non-normal, non-standard or non-factual / detached. (*ibid.* : 331)

Cette étude nous a semblé tout à fait judicieuse car plutôt novatrice et différente des autres travaux consultés jusqu'à présent.

Il semble donc qu'un consensus existe en ce qui concerne les relations entre emphase et fréquence fondamentale, durée et force, depuis Coleman (1914)²⁰, en passant par Trager & Smith (1951 : 52) :

when emphasis is desired on any part of any utterance, several procedures can be used. One can say the whole utterance, or certain parts of it, with

²⁰ Cf. 1.2.3.1.

greatly increased loudness and accompanying extra high, or, in some cases, extra low, pitch.

Enfin pour conclure et résumer cette sous-section, nous citerons Hirst & Di Cristo (1998), dont l'ouvrage passe en revue les systèmes intonatifs de vingt langues avec des études expérimentales et pour chacune une section intitulée « focalisation and contextual effects ». Ils écrivent dans leur introduction :

In the majority of languages described in this volume, focalisation and/or emphasis is said to be best manifested by an extra pitch prominence, giving rise to larger F0 movements often accompanied by extra intensity and duration. (p. 32)

1.2.3.3 Variabilité entre les locuteurs

Dahan & Bernard (1996) ont étudié l'accent emphatique en français et ont réalisé deux expériences : l'une visant à décrire les manifestations acoustiques de l'accent emphatique et à pointer les différences, le cas échéant, entre les locuteurs ; l'autre visant à identifier, parmi les mesures acoustiques, celles qui déterminent la perception de l'emphase. Quatre locuteurs (deux femmes et deux hommes) devant lire vingt-quatre phrases deux fois, une fois sans emphase et une fois avec un accent emphatique sur un mot de la phrase, ont été enregistrés.

Les analyses de la première expérience (durée, fréquence fondamentale et intensité ont été mesurées) révèlent que tous les locuteurs partagent des traits communs : tous ralentissent leur articulation sur les syllabes accentuées et les rendent plus proéminentes en augmentant leur fréquence fondamentale et en baissant l'intensité des syllabes adjacentes. Cependant, ils notent aussi des différences entre les locuteurs (« different prosodic behaviors », p. 367). Pour mettre en relief la différence de longueur, le locuteur 2 raccourcit les éléments non emphatiques, alors que les locuteurs 1 et 3 produisent souvent des pauses après l'élément emphatique. Les autres n'ont pratiquement jamais recours à ce procédé. Le locuteur 4 et le locuteur 3 (mais de façon moindre), placent un pic de la fréquence fondamentale sur la première syllabe du mot-cible alors que les locuteurs 1 et 2 augmentent aussi la F0 sur la première syllabe mais maintiennent le pic sur la dernière syllabe. Ces résultats suggèrent donc que l'accent

emphatique se manifeste par des paramètres communs à tous les locuteurs mais également par des choix spécifiques à chacun :

an emphatic accent involves a common pattern of variations, but also some specific manifestations that depend on speaker choice. (p. 367)

Ainsi, Dahan & Bernard se posent la question du sens pragmatique de l'emphase : les fonctions de l'emphase étant celles d'intensification et de contraste, les différences acoustiques trouvées dans leur étude

might be attributed to cross-speaker differences in intended function of emphasis. (p. 367)

Leur deuxième expérience concerne la perception de l'emphase. Dahan & Bernard ont demandé à des auditeurs de dire s'ils percevaient de l'emphase sur les phrases enregistrées et si oui, à quel endroit. Les scores sont très élevés et les locuteurs réussissent bien à faire passer l'emphase. Là encore les résultats diffèrent selon les locuteurs, certains y arrivant mieux que d'autres. Les locuteurs 1 et 4 ont des scores très élevés alors que le locuteur 3 semble avoir exprimé de l'emphase sans le vouloir et que le locuteur 2 n'a pas toujours réussi à exprimer l'emphase assez clairement.

Dahan & Bernard ont ensuite procédé à une analyse de régression qui montre que l'augmentation de la F0 est le paramètre le plus important dans la perception de l'emphase. D'autres paramètres acoustiques semblent également jouer un rôle dans la perception de l'emphase : la durée des syllabes emphatiques, ainsi que le contraste des durées, et la durée des pauses intervenant avant les accents emphatiques (pas les pauses intervenant après). L'intensité s'avère moins importante. Le rôle du changement d'accent de la dernière syllabe à la première syllabe est révélateur pour la perception de l'emphase en français, mais est d'un intérêt moindre pour l'anglais, le système d'accentuation des mots étant très différent entre les deux langues.

Dahan & Bernard ont choisi comme cible de l'emphase des mots de catégories grammaticales différentes (des noms, des adjectifs et des adverbes) et ils notent de façon tout à fait intéressante que très peu de différences ont été trouvées dans la production et dans la perception de l'emphase de ces différents mots (cf. p. 369).

Cette étude nous semble très pertinente, car elle insiste sur l'importance de prendre en compte les différences entre les locuteurs et les caractéristiques de chacun.

Gussenhoven & Rietveld (1998) ont consacré récemment une étude à ce phénomène et en particulier à la différence de perception que l'on observe entre un locuteur féminin et un locuteur masculin. Ils démontrent grâce à des tests de perception que les auditeurs jugent la proéminence différemment selon que le locuteur est d'un sexe ou de l'autre, alors que les contours de la F0 sont identiques. Les stimulus sont composés de phrases brèves, resynthétisées afin que la voix sonne masculine ou féminine, et dans lesquelles un mot reçoit le contour intonatif H*+L (c'est-à-dire une mise en relief que l'on peut qualifier d'emphatique). Les syllabes accentuées dans les stimulus artificiels de femmes sont jugées moins proéminentes que celles des stimulus des hommes. Cela était un résultat attendu car en effet les locuteurs se fabriquent tous une échelle de référence sur laquelle ils projettent les contours entendus. Une voix de femme a une fréquence fondamentale plus élevée, et la proéminence étant largement perçue en fonction de l'augmentation de la fréquence fondamentale, les résultats n'ont rien d'étonnant :

listeners appear to use the speaker's voice characteristics to estimate the location of the reference scale on which the F0 contour is projected so as to 'read off' the speaker's intended prominence level. (p. 380)

Cette étude confirme donc qu'il est absolument nécessaire d'inclure une composante spécifique à chaque locuteur dans des études perceptives.

1.3 Conclusion

Cet état de l'art nous a permis de constater que l'emphase (ou encore « high involvement » - forte implication, ou « more-than-normal involvement » - implication plus importante qu'à la normale) est un sujet dont on parle beaucoup que l'on définit rarement. L'emphase s'étendant sur de nombreux domaines, il en est souvent plus ou moins question. Le concept est mentionné. Mais peu de travaux traitent de l'emphase à proprement parler. Il nous semble donc d'autant plus intéressant d'étudier ce phénomène plus avant.

On se rend bien compte également en faisant l'état de la question que certains paramètres sont incontournables, et que tous les chercheurs s'accordent à dire qu'ils sont essentiels dans l'analyse de l'emphase. Il s'agit comme nous l'avons vu des paramètres acoustiques suivants :

- la fréquence fondamentale (tous),
- l'intensité (Coleman, 1914 ; Fry, 1955, 1958 ; Selting, 1994 ; Hirst & Di Cristo, 1998) ;
- la durée (Coleman, 1914 ; Erickson & Lehiste, 1995 ; Selting, 1994 ; Hirst & Di Cristo, 1998) ;
- les pauses (Coleman, 1914 ; Tannen, 1984) ;
- le rythme (Coleman, 1914 ; Erickson & Lehiste, 1995 ; Selting, 1994) ;
- la densité des syllabes accentuées (Uhmann, 1992 ; Selting, 1994) ;
- le tilt spectral (Campbell & Beckman, 1995 ; Sluitjer & van Heuter, 1993, 1995) ;
- la qualité de la voix (Tannen, 1984 ; Selting, 1994).

Il s'est également dégagé de ce chapitre qu'il ne faudra pas négliger l'aspect de variabilité entre les différents locuteurs, et donc prendre une base de référence non emphatique pour chaque locuteur que nous étudierons.

Mais la plus grande partie des travaux que nous avons résumés ici est basée sur de la parole de laboratoire, aussi naturelle que possible dans certains cas, certes, mais les

corpus demeurent généralement des phrases artificielles conçues pour l'étude ou des phrases amenées par des questions très précises, donc également artificielles.

Nous espérons que l'intérêt de notre étude résidera principalement dans le fait que nous changeons de base de données et prenons de la parole spontanée. Nous souhaitons innover en quelque sorte pour ne pas nous contenter de refaire sur d'autres stimulus artificiels les analyses déjà effectuées précédemment. Nous étudierons donc les paramètres qui ressortent dans toutes les études décrites dans ce chapitre et tenterons de voir s'ils sont tous aussi pertinents pour la parole spontanée que pour la parole de laboratoire.

De plus nous souhaitons rompre avec les études précédentes en ce qui concerne la méthode. Nous rejoignons l'idée d'une base de données de parole spontanée. En effet, puisque notre corpus n'est pas créé spécialement pour l'étude de l'emphase, il nous faudra trouver une façon appropriée de déterminer quels passages seront considérés emphatiques et ne pas tomber dans le prédéterminé ou le préconçu.

Ces deux points précis (la base de données et la méthode) sont développés dans le chapitre suivant.

Le reste de notre étude est consacré aux analyses instrumentales et là encore la revue que nous venons de faire des études portant sur l'emphase fait bien ressortir deux niveaux : la production et la perception. Ces deux niveaux sont de toute évidence liés, et nous baserons nos expériences sur des tests de perception, et analyserons de quelle façon le son est produit. Les deux niveaux seront donc interdépendants tout au long de notre travail.

Enfin nous pouvons affirmer au vu de toutes les études menées jusqu'ici que l'emphase recouvre d'autres domaines que l'acoustique ou la phonétique et la phonologie. Il faudrait se pencher sur les côtés pragmatiques, lexicaux, syntaxiques, grammaticaux et même gestuels. Nous ne pouvons malheureusement pas, dans le cadre de ce travail, étudier tous ces aspects et nous nous contenterons ici d'une étude phonétique et prosodique.

2 Chapitre 2

Base de données et méthode

Nous nous sommes largement inspirée des études entreprises précédemment pour décider de la méthode à suivre et des expériences à réaliser.

On retrouve dans toutes les études décrites dans le chapitre précédent les mêmes points faibles selon nous :

- la base de données,
- la méthode choisie pour les expériences (par exemple le choix des émotions au départ avec tout ce que cela implique de préconçu).

Nous avons donc essayé d'éviter les mêmes erreurs et nos choix ont été largement influencés par cela.

2.1 *La base de données*

Nous voulions absolument éviter un corpus artificiel, c'est-à-dire des phrases créées par nous et lues par des acteurs par exemple. Nous voulions que notre étude porte essentiellement sur de la parole spontanée, afin de ne pas nous exposer aux mêmes critiques que nous avons formulées nous-même sur les études précédentes et également afin de nous démarquer quelque peu des autres expériences déjà réalisées et pouvoir ensuite comparer nos résultats avec ceux des travaux sur la parole de laboratoire. En effet celle-ci n'est pas sans intérêt, bien entendu, et a permis de réaliser des études très intéressantes et essentielles, les travaux résumés dans le premier chapitre de ce travail en étant la preuve incontestable, mais elle montre ses limites, en particulier dans des domaines comme l'emphase justement ou la focalisation qui sont extrêmement dépendantes du contexte. Beckman (1997 : 12) fait l'apologie du discours spontané pour ces raisons-là :

I have spent so much time motivating why we want to look at spontaneous speech for a better modelling of how focus domain relates to accentuation in English because this is a good example of how lab speech has served us well and of how it fails to serve our needs completely.

Trente ans en arrière, Lehiste (1963 : 353) suggérait déjà que le naturel devait être étudié plus précisément mais en trouvant un équilibre :

... the linguist who wants to use experimental methods in his investigation is forced to trade a certain amount of naturalness for rigorousness in his experimental design... The two requirements appear to occupy the opposite ends of a continuous scale of gradual mutual exclusion. At one extreme, we find complete naturalness coupled with complete lack of control ; at the other end, complete control over the experimental design at the expense of linguistic relevance.

Toute la difficulté d'un travail expérimental fondé sur de la parole spontanée est bien résumée ici et Beckman met également les chercheurs en garde quant à la difficulté de trouver des bases de données de discours spontané analysables (cf. pp.16-20). Il n'est pas évident de faire des enregistrements soi-même.

C'est la raison pour laquelle la recherche d'un corpus nous a pris beaucoup de temps.

2.1.1 Bases de données existantes

Il existe quelques corpus oraux en anglais auxquels nous nous sommes intéressée, mais il n'y en a que très peu :

- le « *British National Corpus* » (BNC) (cf. Clear, 1993 pour les explications), mais il s'agit surtout d'un corpus écrit. Il comporte cependant environ 10% d'oral, essentiellement des enregistrements de cours magistraux, ce qui nous semblait peu attrayant pour un travail sur l'emphase en discours spontané. En effet nous souhaitions travailler sur l'interaction. D'autre part, ce corpus est onéreux (environ 3000 francs français pour une année et 700 francs en ligne) pour le peu de profit que nous aurions pu en tirer.

- Le « *London-Lund Corpus* » (LLC) (cf. Svartvik & Quirk, 1980) regroupe 100 textes oraux de 5.000 mots chacun (ce qui fait un corpus de 500.000 mots d'anglais britannique parlé) mais récoltés dans les années 60/70 pour la plupart, ce qui nous a paru un peu ancien. Ce corpus tire son origine de deux projets : le « *Survey of English Usage* » (SEU), fondé en 1959 par Quirk et qui a servi de base pour la grammaire de Quirk *et al.* (1985) par exemple, et le « *Survey of Spoken English* » (SSE), fondé en 1975 par Svartvik en association avec le SEU (cf. Svartvik, 1990 pour plus de détails). Toutes ces données sont maintenant regroupées sous le nom de « *International Corpus of English – British Component* » (ICE-GB). Il est possible de se procurer ce corpus mais au prix de £300 (3.000 francs français environ). L'avantage de ce corpus est qu'il est entièrement informatisé. Cependant la version orale n'est pas encore achevée.

- Le « *Corpus of Spoken American English* » (CSAE) (cf. Chafe *et al.*, 1991) qui présente l'inconvénient pour nous d'être de l'américain. Nous avons décidé de nous cantonner à l'anglais britannique, celui-ci présentant déjà suffisamment de variétés régionales. Nous reviendrons plus en détails sur ce point²¹.

- Le « *Bergen Corpus of London Teenage Language* » (COLT) (cf. Haslerud & Stenstrom, 1995), qui représente lui aussi une variété d'anglais très marquée, puisqu'il s'agit du parler des jeunes londoniens.

- Le « *Lancaster Spoken English Corpus* » (SEC) semblait très intéressant. Les données ont été recueillies entre 1984 et 1987. Il existe une version informatisée de ce corpus,

²¹ Cf. 2.1.2.

« *MAchine Readable Spoken English Corpus* » (MARSEC) des Universités de Reading et de Leeds (cf. Roach *et al.*, 1993). Ce corpus est déjà entièrement étiqueté du point de vue de l'intonation, ce qui était un point nettement favorable. Peter Roach a eu la gentillesse de nous faire parvenir le corpus MARSEC et nous avons pu l'écouter intégralement. Mais ce corpus uniquement oral est constitué pour sa majeure partie d'enregistrements de bulletins d'informations ou de reportages de la BBC, et également de lectures de passages de romans. Ainsi il ne contient que très peu de passages émotionnels ou emphatiques. L'équipe de Reading et Leeds travaille d'ailleurs en ce moment sur un corpus émotionnel (cf. Roach, 2000).

C'est la tendance actuelle puisque très récemment, plusieurs équipes de recherche se sont mises à travailler à l'élaboration de bases de données orales et émotionnelles, sentant qu'il y avait un manque dans ce domaine, mais elles se heurtent actuellement aux problèmes des droits d'auteur et ne peuvent diffuser leur travail (Douglas-Cowie, Cowie & Schröder, 2000 notamment et Roach, 2000).

Aucun de ces corpus ne nous a paru valable pour la recherche que nous voulions effectuer, mis à part bien entendu les corpus émotionnels mais que nous ne pouvons pas utiliser, et après ces longues recherches infructueuses, nous avons donc décidé de composer nous-même notre propre corpus avec l'inconvénient qu'il ne serait pas étiqueté bien sûr, et surtout que nous serions seule à travailler dessus.

Enfin nous mentionnons le tout nouveau corpus « *IViE : Intonational Variation in English* » dont les principaux instigateurs sont Esther Grabe et Francis Nolan des universités d'Oxford et Cambridge (<<http://www.phon.ox.ac.uk/~esther/ivyweb/>>). Ce corpus a été achevé courant 2001, trop tard donc pour nous, mais il vaut la peine d'être regardé : il contient des données comparables dans neuf dialectes urbains anglais (Belfast English, Bradford Punjabi English, Cambridge, Cardiff, Dublin, London - locuteurs d'origine antillaise -, Leeds, Liverpool, Newcastle) et cinq styles de discours différents (conversation, « map task »²², texte lu et version redite du même texte, phrases contrôlées). Les locuteurs sont au nombre de douze, six garçons et six filles de seize ans. Les enregistrements ont été faits dans des écoles. C'est un corpus qui

²² Le « maptask » est un protocole de recueil de dialogue où l'on fournit à deux personnes une carte d'un lieu fictif en demandant à l'un d'expliquer le chemin de A à B à l'autre. Les deux personnes ne communiquent que par micro sans se voir. Comme il y a de petites différences entre les cartes et que les noms de certains lieux sont choisis pour contenir un maximum de sons sonnants, on obtient une vraie interaction entre les locuteurs avec un petit peu de matière sonore contrôlée.

s'intéresse aux différences régionales et dialectales bien sûr, mais qui peut être utile au chercheur sur d'autres plans. « IViE » regroupe 36 heures de parole. Les auteurs diffusent le corpus à la communauté des chercheurs gratuitement, ce qui est très rare, et cinq heures de parole annotées prosodiquement devraient être prochainement disponibles.

2.1.2 Choix de la variété d'anglais

Avant toute chose, il nous fallait décider quelle variété d'anglais britannique choisir (nous savions déjà que nous ne voulions pas travailler sur l'anglais américain). Ce qui vient tout de suite à l'esprit est l'anglais britannique standard, ce que l'on appelle aussi Received Pronunciation, BBC English (terme qui perd de son sens, les personnes travaillant à la BBC n'ayant plus systématiquement, depuis quelques années, un anglais standard) ou encore Queen's English.

Pourtant, nous pensons que la majorité des britanniques ne parlant pas l'anglais standard (loin s'en faut), il n'était pas forcément représentatif de choisir un corpus comportant uniquement de l'anglais standard. Nous souhaitions avoir différents accents régionaux, pas trop prononcés néanmoins.

La variété d'anglais a donc pesé lourdement dans nos choix, comme nous allons l'expliquer tout au long de la section suivante.

2.1.3 Constitution de la base de données

Comment constituer une base de données ? Partir de quoi ? Télévision, radio, cinéma, enregistrements spontanés ? Telles étaient les questions qui nous tourmentaient.

Nous avons commencé à travailler sur l'enregistrement d'un débat car l'un de nos collègues travaillait sur cette base de données (Rouveyrol, doctorat en cours). Il s'agit de l'émission télévisée *Question Time* qui avait lieu ce jour-là à Belfast en Irlande du

Nord²³. Le débat a lieu entre plusieurs personnalités politiques britanniques (environ 12500 mots et un peu plus de 47 minutes de son). Cela nous a semblé bien plus pertinent que tous les corpus auxquels nous nous étions intéressée auparavant car il y a des discussions, les gens s'interpellent, tentent de convaincre ou de se défendre, etc... Nous avons donc plus de chance de trouver de l'emphase. Il s'agit bien entendu d'un type de discours particulier, le discours politique, mais cela n'en reste pas moins de la parole spontanée naturelle.

L'émission se passe à Belfast et l'on s'attend donc à entendre de l'anglais d'Irlande du Nord où l'accent régional est très fort, d'autant que l'intonation diffère de l'intonation de l'anglais standard (comme c'est le cas pour l'anglais d'Ecosse ou du Pays de Galles). Mais les personnes présentes à ce débat sont des personnalités politiques et elles possèdent donc un anglais qui se rapproche de l'anglais standard. Leur accent n'est vraiment pas très marqué. Par contre, le public pose des questions dans cette émission et certains passages sont très « irlandais ». Nous n'avons pas choisi ces passages-là. Nous n'avons pas sélectionné pour notre corpus primaire les morceaux où l'accent était trop fort.

Comme nous ne pouvions pas travailler uniquement sur du discours politique, de peur de se voir reprocher, fort légitimement, que ce n'est pas du tout représentatif de la parole spontanée naturelle, nous avons continué à chercher d'autres bases de données. Nous avons pensé qu'après la télévision il serait bon de voir ce qui se passait à la radio. Nous avons donc fait des enregistrements de BBC World Service (que nous captions de France) et nous avons sélectionné dans ces enregistrements une discussion entre deux femmes britanniques d'origine asiatique sur les mariages forcés et arrangés²⁴. Hormis le type d'émission (nous avons déjà précisé que nous voulions travailler sur l'interaction, il nous fallait donc une émission où plusieurs personnes parlent et échangent des idées), ce sont les différents accents qui ont motivé notre choix une fois encore. La BBC World Service, comme son nom l'indique, s'adresse à des auditeurs de tous les pays et les personnes invitées à parler sont rarement britanniques et ont alors un accent marqué de leur pays d'origine, même si leur anglais est très bon.

²³ L'enregistrement intégral se trouve sur le CD ROM : *Corpus : Son : QuestionTime*.

²⁴ L'enregistrement intégral se trouve sur le CD ROM : *Corpus : Son : BBC*.

Les deux femmes qui s'expriment dans le passage que nous avons sélectionné sont d'origine asiatique mais sont britanniques, elles ont toujours vécu en Grande Bretagne et ont un anglais de « native speakers ». L'une d'entre elles ne parle pas une variété d'anglais tout à fait standard, mais cela ne nous importait pas, puisque son accent n'était pas trop fort, comme nous l'avons déjà dit plus haut. Les deux femmes ne sont pas du même avis dans cet enregistrement, et c'est un sujet qui leur tient beaucoup à cœur, de très jeunes filles étant impliquées, ce qui provoque des passages émotionnels et emphatiques. Le passage est assez court (environ 1150 mots et environ 7 minutes de son).

Comme nous avons deux passages où les gens s'opposent, nous avons pensé qu'il serait intéressant d'étudier l'emphase dans un passage où les différents locuteurs ne sont pas forcément en désaccord. De plus, nous souhaitons que notre corpus comprenne de la conversation de tous les jours (« everyday conversation »). C'est un des points auquel nous tenions tout particulièrement car très peu d'études sont faites sur ce type de discours, alors qu'il s'agit tout de même du plus courant. Pour ce faire, il nous fallait enregistrer des anglophones ayant une discussion tout à fait banale. Mais cela est très difficile à réaliser dans des conditions naturelles à cause du bruit de fond. Il nous fallait un enregistrement de qualité afin de pouvoir l'analyser et procéder à des manipulations. Nous ne pouvions pas par exemple enregistrer des gens dans un bar autour d'un café et il nous semblait délicat d'aller placer un micro chez les gens. Nous avons donc demandé à quatre anglophones de bien vouloir avoir une discussion banale, sur tout et n'importe quoi, dans un studio d'enregistrement. Pour que les choses paraissent le plus naturel possible, nous leur avons demandé de faire comme s'ils buvaient un café dans un bar (et ils buvaient effectivement un café !). Nous les avons enregistré pendant 20 minutes (21 minutes 44 exactement et environ 5000 mots)²⁵. Dans ce genre d'enregistrements, le plus gros problème est celui du chevauchement. En effet, dans une conversation naturelle, les gens parlent beaucoup les uns en même temps que les autres, surtout lorsqu'ils sont nombreux. Sur ce point, le fait d'avoir quatre locuteurs augmentait certainement le chevauchement, mais cela avait par ailleurs l'avantage que nous avions quatre personnes différentes et l'emphase varie peut-être d'une personne à l'autre.

²⁵ L'enregistrement intégral se trouve sur le CD ROM : *Corpus : Son : Conversation*.

Parmi ces quatre personnes, nous avons deux femmes, l'une de 48 ans possédant une variété d'anglais du Sud de l'Angleterre (standard), l'autre de 25 ans, originaire des Midlands, possédant un accent moins standard donc, et deux hommes, tous deux de Londres, l'un de 30 ans et l'autre de 24 ans. Nous avons donc quatre personnes assez différentes, ce qui nous a paru très intéressant, leur anglais n'étant pas tout à fait similaire mais pas trop marqué. Pour en revenir au problème du chevauchement, nous avons tout simplement supprimé pour notre corpus primaire les passages où deux personnes ou plus parlent en même temps car il aurait été trop difficile d'en faire l'analyse ensuite.

Nous avons donc des extraits de télévision, des extraits de radio, et de la conversation de tous les jours, et nous avons pensé qu'il serait intéressant de prendre un passage de film récent. Nous avons pourtant expliqué que nous ne voulions pas tomber dans le piège de phrases lues par des acteurs. Mais dans un film, les acteurs tentent de recréer des situations naturelles. Il est vrai que les choses sont parfois exagérées au cinéma, et c'est justement cela qui est peut-être pertinent pour l'analyse de l'emphase. Si un acteur veut rendre une phrase ou un mot ou un passage emphatique, il va utiliser un procédé qui est normalement naturel et spontané, voire inconscient, de façon consciente précisément et en l'augmentant pour être sûr que l'emphase soit bien claire. Nous avons donc estimé qu'il serait bon de prendre un bout de film pour notre analyse, afin de rendre notre corpus varié et assez complet.

Nous avons malheureusement manqué de temps pour faire l'analyse de quelques morceaux de film et ce travail ne contient donc rien dessus, mais nous ne perdons pas l'espoir d'étudier par la suite des dialogues de film²⁶.

Ainsi notre base de données était constituée (sans le film donc), base de données d'environ 18500 mots et représentant environ une heure et quart de son, avec des locuteurs très différents, des variétés d'anglais diverses également mais pas trop

²⁶ Nous avons déjà sélectionné le film *Peter's Friends* car nous voulions un film récent se situant à notre époque pour entendre de l'anglais contemporain des années 80 ou 90. Ici, une fois de plus, le problème de la variété d'anglais s'était posé. Le cinéma britannique de ces dernières années est un cinéma très réaliste, dont le meilleur représentant est probablement le réalisateur Ken Loach. Mais ses films ont souvent pour cadre le Nord de l'Angleterre et la variété d'anglais est très marquée, trop marquée malheureusement, nous a-t-il semblé.

marquées, trois sources disparates, et toujours des situations d'interaction. Cette base de données constitue ce que nous appellerons notre corpus secondaire et nous appellerons corpus primaire une sélection de passages tirés de ce corpus secondaire sur lesquels nous avons travaillé plus précisément. Ces passages sont les morceaux sélectionnés pour nos tests de perception (cf 2.2.3).

Nous avons obtenu la transcription en anglais de l'émission *Question Time*, et avons transcrit orthographiquement nous-même toutes les autres données²⁷.

²⁷ La transcription orthographique des trois passages qui constituent notre base de données se trouve sur le CD ROM : *Corpus : Transc orth : Question Time* ; ainsi que *Corpus : Transc orth : Conversation* ; et enfin *Corpus : Transc orth : BBC*.

2.2 Méthode

Afin de ne pas tomber dans le préconçu et l'expérience trop directive, il nous fallait trouver un moyen objectif pour décider quels passages étaient emphatiques et allaient nous être utiles.

La seule solution semblait être de demander l'avis d'anglophones, de « native speakers ».

Nous avons donc sollicité les anglophones que nous connaissions pour qu'ils écoutent des morceaux de notre corpus et nous indiquent les passages qui leur semblaient emphatiques. Nous avons pensé qu'il nous faudrait au minimum quinze personnes d'âge et de sexe différents. La solution de facilité était bien sûr de s'adresser aux personnes avec lesquelles nous travaillons dans le département d'anglais de l'Université de Provence. Il s'agit donc essentiellement d'enseignants, de lecteurs et d'étudiants du département²⁸. Un seul auditeur est spécialiste de phonologie. Nos auditeurs sont donc considérés comme auditeurs naïfs.

Cela constitue le point de départ indispensable de tout notre travail.

Cependant un problème se pose pour la mise en place du test. Comment demander aux gens de noter l'emphase ? Fallait-il leur demander de dire si les passages étaient emphatiques ou non, ou bien leur donner une échelle avec plusieurs degrés d'emphase à évaluer ? Cela revient à se poser la question de savoir si l'emphase est un phénomène catégoriel ou continu.

2.2.1 L'emphase : phénomène catégoriel ou continu ?

C'est la question que se sont posés Ladd & Morton (1997). Ils tiennent pour acquis le fait que l'emphase est rendue par les variations mélodiques et donc la fréquence

²⁸ Cf. CD ROM : *Sondages : auditeurs*, ou annexe 2a.

fondamentale²⁹. Ce qui les intéresse, c'est de savoir si l'emphase est continue ou catégorielle. Existe-t-il une échelle allant du moins emphatique au plus emphatique, sur laquelle la frontière entre non-emphatique et emphatique serait floue, et donc cette partie de l'échelle serait non utilisée afin d'éviter l'ambiguïté, ou bien l'emphase serait-elle un phénomène catégoriel, c'est-à-dire qu'il n'y aurait pas d'échelle, un énoncé serait soit emphatique soit non emphatique ? Il semblerait logique, à première vue, sans aller trop loin, de dire que l'emphase est certainement une question de degrés et est donc continue. Si l'on prend une phrase que l'on prononce plusieurs fois avec le même contour intonatif et en faisant varier la hauteur mélodique, on arrive à des degrés d'emphase différents. Plus la hauteur mélodique est grande, plus l'emphase est forte :

it is trivial to observe that when you raise your voice your utterance sounds more emphatic, and also – this is the gradient part – that the more you raise your voice, the more emphatic it sounds. (Ladd, 1994 : 43).

Pourtant, il semble que ce ne soit pas si simple.

L'étude de Gussenhoven & Rietveld (1988)³⁰ les a fait réfléchir sur ce problème car elle montre que la fréquence fondamentale agit sur la phrase tout entière pour signaler de la proéminence et non pas sur chaque accent individuel. Ladd *et al.* (1994)³¹ vont plus loin et pensent qu'il existe au premier plan une échelle catégorielle, avec des énoncés normaux et des énoncés emphatiques, et au deuxième plan, à l'intérieur des énoncés emphatiques, une échelle continue, avec différents degrés d'emphase. Les travaux de Hirschberg & Ward (1992)³² tendent à prouver au contraire que l'emphase serait catégorielle si l'on considère comme Ladd & Morton (1997) que le contour L*+HL-H% avec l'interprétation d'incrédulité est un contour emphatique (« higher degree of involvement »), alors que quand l'interprétation est l'incertitude, il s'agit d'un contour « normal ». Néanmoins, Hirschberg & Ward n'ayant utilisé dans leur étude que deux hauteurs mélodiques extrêmes (une basse et une élevée), on ne peut conclure à l'existence d'une frontière claire entre énoncés non emphatiques et énoncés emphatiques.

²⁹ Cf. 1.2.3.2. pour le consensus qui existe au niveau des paramètres acoustiques liés à l'emphase.

³⁰ Etude décrite en 1.2.2.3.

³¹ Cf. 1.2.2.3.

³² Travaux décrits en 1.2.1.4.

Morton (1993) a alors tenté de savoir si cette frontière existait. Elle a manipulé la hauteur mélodique d'une phrase (*It's Diana again*) avec l'accent sur le dernier mot et a demandé à des auditeurs de commenter leur perception du pic mélodique. Dans certains cas ils avaient le choix entre emphatique et non emphatique, et dans d'autres cas ils disposaient d'une échelle de 0 à 10 et devaient placer le degré d'emphase perçue. Les résultats obtenus donnent une prépondérance des jugements non emphatiques en bas de l'échelle et un changement soudain à une prépondérance des jugements emphatiques sur le haut de l'échelle. Ils tendent donc vers l'hypothèse catégorielle.

Mais Ladd & Morton (1997) ont souhaité continuer la recherche pour arriver à des résultats plus convaincants. Ils ont donc réalisé quatre expériences. Pour ces travaux, des phrases courtes avec un seul pic accentuel et un schéma intonatif de type « fall-rise », c'est-à-dire ton creusé (Faure, 1962), ont été enregistrées, dans une version normale puis emphatique. Ces phrases ont été resynthétisées et modifiées pour les besoins des expériences.

La première étape consiste en un prolongement de l'étude de Morton (1993) afin de confirmer les résultats obtenus et d'identifier les variables susceptibles d'avoir une influence dans les discontinuités des courbes de réponse de la tâche d'évaluation. Les variables les plus pertinentes étaient la F0 (qui fait la distinction entre normal et emphatique) et l'instruction. Deux instructions différentes ont été choisies. L'une est 'paralinguistique' : on demande aux auditeurs de noter le degré d'emphase sur une échelle de 0 à 10 ; l'autre est 'linguistique' : on demande aux auditeurs de choisir entre deux contextes dans lesquels la phrase aurait pu être dite (l'un induisant de l'emphase et l'autre non) et d'en juger la probabilité sur une échelle de 0 à 10. Les résultats montrent que les signes acoustiques de l'emphase survivent aux manipulations de la F0 et que la nature de l'instruction joue un rôle important, l'instruction paralinguistique induisant des jugements plus emphatiques que l'instruction linguistique (Ladd & Morton, 1997 : 324).

La deuxième expérience vise à découvrir si la distinction normal/emphatique est comparable à d'autres différences phonologiques telles que le lieu d'articulation ou la distinction voisé/non voisé pour les consonnes plosives. La méthode de perception catégorielle classique a été utilisée sur une seule phrase (choisie en fonction des

résultats du premier test) non emphatique et avec l'instruction linguistique. Onze stimulus en sont dérivés. Deux phases sont nécessaires :

- une phase d'identification où les auditeurs doivent choisir entre « everyday occurrence » (énoncé neutre) et « unusual experience » (expérience inhabituelle) ;
- une phase de discrimination pendant laquelle les auditeurs entendent deux phrases à la suite et doivent dire si elles sont identiques ou différentes.

Les résultats montrent que les auditeurs sont capables de distinguer deux phrases dont les hauteurs mélodiques diffèrent très peu, et ce à n'importe quel niveau sur un continuum entre non emphatique et très emphatique. Pourtant ceci n'est vrai que pour les paires de phrases dans lesquelles la première a l'accent mélodique le plus bas. Si c'est la deuxième phrase qui a l'accent mélodique le plus bas, alors les auditeurs ne font plus la différence entre les deux phrases. Aucune explication n'est donnée si ce n'est peut-être l'effet de déclinaison³³, c'est-à-dire qu'à cause de ce phénomène auquel l'auditeur est habitué et qu'il anticipe, le second pic accentuel dans une phrase est perçu aussi proéminent que le premier quand il est en fait un petit peu plus bas.

La troisième expérience a pour but de trouver les facteurs qui font que la frontière entre les catégories perceptuelles pourrait changer. C'est l'alignement du pic mélodique qui est le plus pertinent et qui est étudié ici. Il est montré que plus le pic arrive tard, plus l'emphase est perçue. La méthode de perception catégorielle semble montrer ses limites. En effet, si on ne propose que deux choix aux auditeurs, ceux-ci sont tentés de choisir entre les deux, même si ce n'est pas leur intuition de départ. Les résultats sont donc faussés.

Dans la quatrième expérience, les auditeurs entendent encore des paires de phrases mais séparées par de plus longs moments. La perception en est modifiée et le nombre de jugements emphatiques augmente. Le même résultat que dans l'expérience 2 apparaît pour l'ordre des paires et n'est pas constant sur le continuum des valeurs de F0 :

the perceived similarity of two single-accent utterances differing only in pitch excursion depends quite substantially on the order in which the two are presented and, more importantly, this order effect is not constant throughout the continuum of pitch range. (Ladd & Morton, 1997 : 339).

³³ Cf. 1.2.3.2 et Pierrehumbert, 1979 ; Terken, 1991 ; Gussenhoven & Rietveld, 1988 ; Cooper *et al.*, 1985.

Les auteurs n'aboutissent finalement pas à la conclusion que l'emphase est un phénomène catégoriel, ni vraiment continu. Ils avancent l'idée que si l'on ne peut prétendre que l'emphase est perçue de façon catégorielle, il peut par contre être utile de dire que l'emphase est interprétée de façon catégorielle³⁴.

Ils ajoutent que d'autres paramètres acoustiques et pragmatiques jouent un rôle important dans l'interprétation, dont la F0, la qualité de la voix, le contenu lexical, le contexte, les relations entre le locuteur et l'auditeur, etc...

Cette intéressante étude, très poussée et très complexe, invite les futures recherches à ne pas négliger les effets de l'ordre des accents et des stimulus, et à ne pas travailler au niveau de la syllabe accentuée uniquement.

En ce qui concerne notre étude, la question de savoir de quelle manière traiter l'emphase et la faire juger par des auditeurs n'est pas résolue. Il était cependant indispensable de mettre ce problème au jour.

L'étude de Ladd & Morton (1997) met également en relief le problème de l'instruction donnée dans un test de perception. Ce point nous semble essentiel.

2.2.2 Instructions et présentation

Une étude de Streefkerk, Pols & 'ten Bosch (1997) met justement en évidence les problèmes liés à la présentation et aux instructions données dans les tests de perception de la proéminence ou encore emphase :

Naïve listeners are asked to mark those words or syllables spoken with emphasis (this is an operational definition of prominence). (p. 101)

Ils montrent que les résultats varient selon les instructions données. Les phrases utilisées sont des phrases lues. Les trois chercheurs mettent en place deux tests où la présentation acoustique des phrases est identique mais les instructions et la présentation différentes. Dans le premier test, les auditeurs voient les phrases sur un moniteur écrites normalement, et ils doivent marquer les mots mis en relief. Dans le deuxième test, les

³⁴ « While we cannot legitimately speak of pitch range distinctions being categorically perceived, it may be useful to think of them as being categorically interpreted » (Ladd & Morton, 1997 : 339).

phrases sont écrites avec un tiret et un blanc séparant chaque syllabe, et les auditeurs doivent marquer les syllabes mises en relief.

Dans l'expérience de perception des syllabes, les sujets utilisent un mode de perception plus analytique en comparant chaque syllabe avec celles qui l'entourent. Dans le test de perception des mots, l'attention des sujets est attirée à un niveau supérieur, le niveau de la phrase ou du syntagme. Les auditeurs comparent chaque mot avec les autres mots de la phrase. Il est donc logique que dans l'expérience sur les mots, les scores soient moins élevés (2,9 en moyenne pour les mots et 5,1 pour les syllabes). Moins de mots sont marqués proéminents que de syllabes car les auditeurs comparent moins de mots entre eux que de syllabes entre elles.

Cette étude nous incite donc à réfléchir sur les consignes à donner dans de tels tests. Streefkerk *et al.* (1997) favorisent apparemment le test de perception sur les mots car ils réalisent une troisième expérience de perception pour tester l'importance de la fréquence fondamentale et les consignes portent sur le mot. Cependant, si nous sommes persuadée qu'il y a en effet une différence importante selon l'instruction, il nous semble dommage de réduire l'expérience de perception à l'une ou l'autre des deux catégories. Streefkerk *et al.* ne mentionnent pas un troisième niveau qui pourrait être celui d'un passage entier marqué emphatique. Ne serait-il pas intéressant d'imaginer que la proéminence d'un mot ou même peut-être d'une syllabe pourrait faire percevoir tout un passage emphatique ?

D'autre part, le problème de la présentation est soulevé dans l'article de Streefkerk *et al.*, mais de la présentation écrite. Dans leurs expériences, Streefkerk et ses collaborateurs font écouter des phrases lues, hors contexte, trois fois de suite. Il est vrai que leur étude porte sur la perception de la proéminence et non de l'emphase au sens où nous l'entendons ici, à savoir une implication plus importante de la part du locuteur³⁵, mais il nous semble que la perception risque d'être faussée quand des phrases sont données hors contexte³⁶. En effet ce phénomène ne se produit jamais dans la vie de tous les jours et nous interprétons tout discours et donc toute emphase en fonction du contexte et de la situation d'énonciation.

³⁵ Cf. 1.1.5.

³⁶ Cette remarque nous a d'ailleurs été faite par un des auditeurs de nos sondages (cf. 2.2.4.1.).

Ces considérations sont essentielles à nos yeux et rejoignent la discussion tenue à propos du choix du corpus, mais importent également pour le choix des tests de perception qui sont incontournables et surtout le choix des instructions et de la présentation. Cela nous amène à nous demander s'il ne serait pas justement intéressant de réaliser deux expériences comparatives : l'une où les phrases sont données avec le contexte, et l'autre où les phrases sont données hors contexte. On pourrait ainsi voir si les résultats sont les mêmes ou s'ils diffèrent de façon importante et justifier ou infirmer de façon scientifique (ou tout au moins empirique) notre argument en défaveur des stimulus hors contexte³⁷.

2.2.3 Tests perceptuels

2.2.3.1 Méthode

Nos problèmes de base ne trouvant pas de solutions dans la littérature, nous nous sommes dit qu'il fallait faire des essais et nous avons débuté par un test pilote (que nous appellerons « sondage ») tout simple qui consistait à donner une cassette et un fascicule avec le texte de la cassette (écrit en anglais normal et sans ponctuation pour ne pas influencer l'écoute) à chaque auditeur, qui pouvait ainsi l'emporter chez lui et l'écouter quand bon lui semblerait³⁸.

Nous avons commencé par l'enregistrement du débat télévisé. Pour que ce sondage ne soit pas trop long pour nos auditeurs et pas trop difficile, nous avons découpé l'enregistrement en plusieurs passages, sélectionnés de façon tout à fait arbitraire par nous-même. Nous avons choisi les passages qui nous semblaient les plus emphatiques, mais également des passages qui nous semblaient dénués de toute emphase. Ces passages de trois ou quatre lignes sont répétés trois fois et l'auditeur doit souligner le ou les mots qui lui semblent emphatiques. Nous n'avons volontairement pas donné de définition trop précise de l'emphase afin de ne pas influencer nos auditeurs et nous

³⁷ Cf. 5.2.1, où nous réalisons cette comparaison.

³⁸ Les sondages sont disponibles sur le CD ROM : *Sondages* : *Son* : *Sondage1.wav* ; *Sondage2.wav* ; *Sondage3.wav*, et *Sondages* : *Instructions*. Les instructions sont également insérées en annexe 2b.

avons utilisé la définition qui ressort du chapitre précédent³⁹. Les instructions sont les suivantes :

What you have to do is listen to a TV debate and underline all the passages or words which you feel are EMPHATIC. What I mean when I say emphatic is NOT NEUTRAL, with a special INVOLVEMENT on the part of the speaker. For example, ask yourselves if you would utter the same sentence in the same way in a neutral context. What I want you to tell me is what you think, what you feel is being made prominent in some way.

La première version de ce sondage était mauvaise car nous nous sommes rendu compte grâce à la première personne testée que les passages que nous avons pris étaient trop longs, et que l'auditeur n'avait pas bien le temps de souligner et était du fait fort troublé, devant aller trop vite. Nous avons donc, sur ses conseils, découpé le texte différemment : nous avons enregistré des passages un peu longs deux fois, puis choisi à l'intérieur de ces passages des séquences très courtes (une seule phrase ou même parfois une seule unité intonative) répétées deux fois également.

Nous avons demandé à nos auditeurs de nous faire tous les commentaires qui leur semblaient utiles afin d'améliorer les tests ultérieurs :

I have left some space for you to make comments which might be useful to me. For instance you can tell me if you hesitated or had doubts or felt the emphasis was less strong or if you thought it was ambiguous or whatever might help me. [...] I would also be grateful for any suggestions, remarks, ideas, etc. which might help me progress.

Ce premier sondage nous a permis d'affirmer qu'il existe plusieurs degrés d'emphase. En effet certains auditeurs ont trouvé nécessaire de distinguer différents degrés d'emphase, certains en donnant deux (« weak » et « strong »), d'autres trois (« weak », « strong » et « very strong » ou bien « slight », « medium » et « strong »).

Grâce à cela, nous avons modifié les consignes de nos second et troisième sondages, qui portaient respectivement sur la conversation entre nos quatre anglais et le passage de la BBC World Service. Nous avons fait pratiquement la même chose avec cette différence que nous n'avons enregistré les passages longs qu'une fois (ils étaient enregistrés deux fois dans le sondage 1), et nous avons demandé aux auditeurs de distinguer au moins deux degrés d'emphase, « slight » et « strong », trois s'ils en ressentaient le besoin

³⁹ Cf. 1.3.

(« medium » en plus). Nous avons rajouté aussi le niveau de la syllabe dans nos instructions car dans le premier sondage, certains auditeurs nous avaient parfois souligné une seule syllabe d'un mot. Ainsi nous leur avons demandé de souligner les syllabes, mots ou passages qui leur semblaient emphatiques. Cela rejoint naturellement le problème du choix de l'instruction et du classement de l'emphase en phénomène catégoriel ou continu.

Ces sondages devaient nous permettre de mieux déterminer la méthode à suivre et les passages à choisir pour le véritable test de perception qui serait la base de nos travaux. Nous avons rentré dans une feuille de calcul Excel les résultats de ces sondages en numérotant chaque passage, chaque mot à l'intérieur de chaque passage et chaque syllabe à l'intérieur des mots pour lesquels le découpage en syllabes était nécessaire d'après nos résultats. Ceci nous permettrait de faire des statistiques et de déterminer un indice d'emphase⁴⁰.

C'est au vu des résultats de ces sondages, grâce aux expériences faites précédemment, mais aussi grâce aux commentaires que certains de nos auditeurs ont bien voulu nous faire, que nous devons diriger nos choix pour le test final et pour le choix des paramètres à analyser une fois le test de perception fait.

2.2.3.2 Commentaires

Il nous faudra distinguer, comme nous l'avons déjà dit, différents degrés d'emphase. Il est intéressant de noter que pour le sondage 1, certains auditeurs ont d'eux-mêmes distingué plusieurs degrés d'emphase sans que nous leur ayons demandé et que dans le sondage 2 où nous souhaitions qu'ils nous donnent deux ou trois degrés d'emphase, certains ont ressenti le besoin de noter dans les emphases déjà classées fortes, une force supplémentaire. L'auditeur DT⁴¹ par exemple soulignait les mots emphatiques et notait L (« light »), M (« medium ») et C (« clear »), et a parfois marqué à côté du C un point d'exclamation qui semble symboliser un degré encore plus élevé.

⁴⁰ Cf 2.2.3.3.

⁴¹ Les auditeurs et les locuteurs sont tous nommés par leurs initiales, cf. annexe 2a.

- Pratiquement tous les auditeurs sont d'accord pour dire qu'il s'agit d'un exercice difficile et presque tous ont beaucoup hésité et ont eu du mal à dire sur certains passages s'il s'agissait d'emphase ou simplement d'hésitation de la part du locuteur. Ceci nous amène à penser que le rôle des pauses est très important.

- Certaines personnes nous ont dit que le sens des mots était essentiel. Ils ont précisé par exemple que deux mots étaient liés : sur un mot DV note que le locuteur met l'emphase pour que l'on voie bien la comparaison qu'il est sur le point de faire : « emphasizes so we see the comparison he's about to make » et il ajoute sur le mot suivant que l'emphase sert à mieux établir la connexion entre les deux mots pour l'auditeur : « and here is true comparison. The two are linked and the listener can notice true connection more easily through emphasis. ». DV note encore par exemple que le mot « employment » contient une emphase sur la première syllabe car il est en opposition avec « unemployment » qui suit juste après (c'est lui qui souligne).

- RT est également sensible au contexte et en particulier à la redondance. Il lui paraît peu probable qu'un mot déjà donné soit mis en relief et pourtant c'est l'impression qu'il a sur le mot « snow » : « sounds like emphasis but unlikely as the term has already been mentioned here. ». Cela signifie que les personnes qui écoutent sont sensibles à la façon dont est dit le mot (il dit « sounds »), peut-être l'intonation ou l'intensité ou l'allongement de la voyelle, mais aussi au contexte et aux notions d'information nouvelle et déjà donnée. DV note à un moment donné dans le sondage 2 qu'un mot est emphatique car il s'agit d'un élément nouveau : « introduction of new factor ». PP insiste sur ces notions (information nouvelle ou ancienne) aussi dans le commentaire qu'il fait suite au sondage 1. Il explique qu'il est souvent difficile de savoir si un mot est emphatique ou non et que c'est bien souvent à ces notions-là qu'il a recours pour trancher. MH, quant à lui, soulève à propos du contexte un autre problème : il trouve que les phrases prises séparément et sorties de leur contexte sont tout à fait différentes des mêmes phrases entendues au préalable en contexte dans le passage long. Selon lui, le fait d'entendre les phrases à nouveau deux fois et sorties de leur contexte fausse considérablement la perception de l'emphase. Cette remarque est très perturbante et à la

fois fortement intéressante. Elle remet en question les tests de perception classiques dans lesquels les phrases sont justement hors contexte. Il est cependant le seul sur les 18 auditeurs du sondage 1 à nous faire cette remarque.

- De nombreuses personnes estiment aussi que la façon de s'exprimer est déterminante. Dans le sondage 1, PM a la façon de parler la plus emphatique selon JH alors que CR trouve que dans un des passages il ne met rien en relief et que c'est peut-être justement là une façon d'être emphatique: « one way of being emphatic is to articulate a whole phrase (or more) without putting emphasis on any word at all like here » (c'est lui qui souligne). TR précise que JT a une façon de mettre en relief tous les mots. La plupart des auditeurs sont d'accord pour dire que MM a une façon de parler tout à fait particulière : « way of speaking tends to obscure some terms which would otherwise be emphatic » ou encore « unusually flat, non marked diction ». GP pense que « it has a lot to do with the character and nationality of a person ». En résumé, il semble qu'il sera nécessaire de ne pas oublier que chaque locuteur est unique et a un fonctionnement propre, et que nous devons y penser lors de l'analyse des résultats, afin de comparer le locuteur avec lui-même et pas seulement avec les autres locuteurs. Cela confirme les études faites sur la variabilité entre locuteurs⁴².

- Le rythme et les pauses semblent jouer un rôle également. RT note : « most emphatic utterances here due to intonation, last one 'three' due to pause and slower speech ».

- DV nous fait une remarque très intéressante sur l'interaction : dans la phrase « Did you do that last year as well ? », il pense que « as well » est mis en relief par JC pour montrer qu'il écoute VB : « I think JC emphasizes 'as well' to show he's listening to VB, showing interaction ». Il est passionnant de se dire que l'emphase joue peut-être parfois un rôle à l'intérieur du cadre interactionnel et cela ne peut que renforcer notre conviction de départ, à savoir qu'il est absolument essentiel d'étudier du discours et non pas uniquement de la parole de laboratoire. DV parle plusieurs fois d'emphase en relation avec des passages ou des répliques humoristiques dans la conversation du sondage 2. Dans la réplique de JC « I'd like to try a heat wave », il pense que JC met

⁴² Cf. 1.2.3.3.

« try » en relief pour dire qu'il n'y a jamais de vagues de chaleur en Angleterre et cela n'est possible qu'en relation avec les phrases précédentes et le contexte évidemment puisqu'ils viennent de parler de la chaleur en France par rapport au temps pluvieux de l'Angleterre. C'est la même chose un peu plus loin à propos d'une histoire de cérémonies officielles et de chapeaux. L'emphase est mise sur le mot chapeau pour rendre la phrase amusante. Il est aussi question de surprise ou de réactions émotionnelles dans les commentaires de DV tout au long de la conversation. Encore une fois l'accent est mis sur le contexte et la relation d'interaction. De tels phénomènes sont impossibles à trouver en parole de laboratoire.

2.2.3.3 Résultats des sondages et test de perception

Nous avons rentré dans une feuille de calcul Excel les résultats des sondages⁴³, c'est-à-dire que chaque mot est rentré en ordonnées et les auditeurs en abscisses avec les résultats par mot. Si le mot n'est pas marqué emphatique, le chiffre correspondant est 0. Quand un degré est exprimé (sondages 2 et 3), nous avons noté 1 (degré faible), 2 (degré moyen), ou 3 (degré élevé) en face du mot, et quand aucun degré n'est donné (sondage 1 uniquement) nous avons pris le chiffre 3.

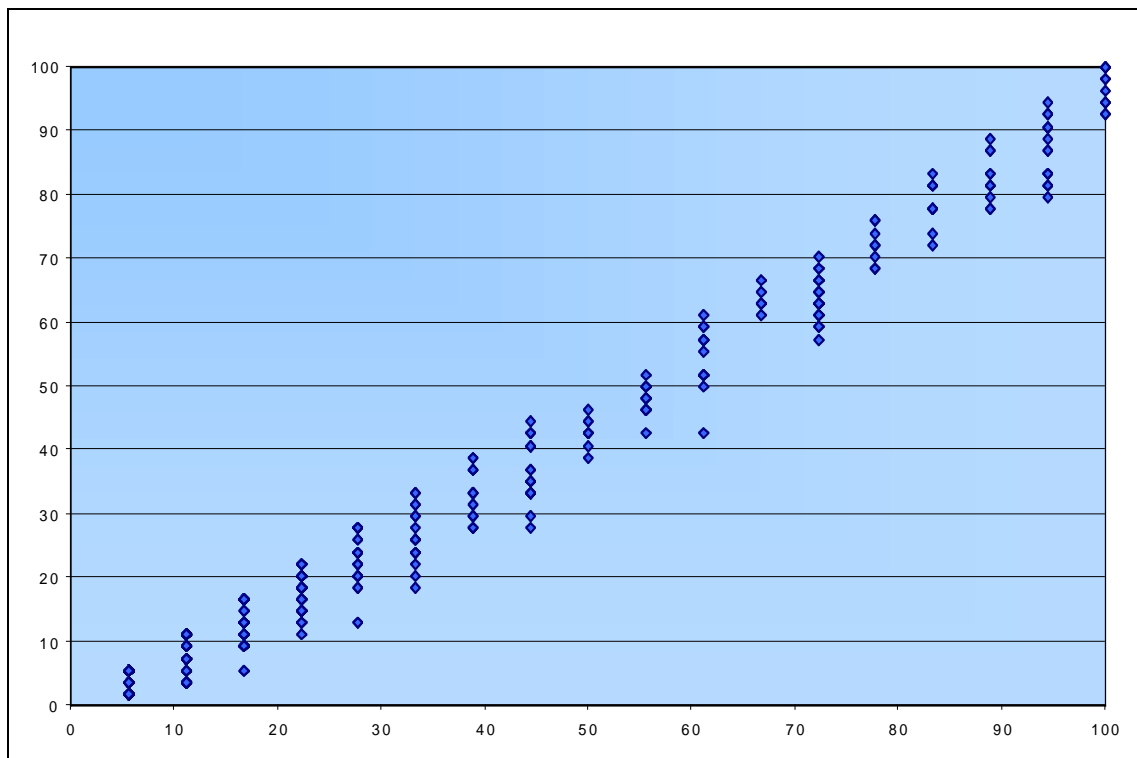
Ceci fausse sûrement les résultats du sondage 1 car certains auditeurs n'ont donné aucun degré d'emphase, donc les résultats sont toujours 3, alors que dans le même sondage, certains en ont donné, et les résultats vont de 1 à 3.

A partir de ces résultats, nous avons déterminé un pourcentage d'emphase : la somme des auditeurs qui ont donné un mot emphatique, sans prendre en compte le degré, multiplié par 100 et divisé par le nombre d'auditeurs. Ce pourcentage (qui ne prend donc pas en compte l'indice du degré) donne déjà une notion de degré. Si un mot est marqué emphatique par 75% des auditeurs, on peut dire qu'il est plus emphatique qu'un mot marqué par seulement 20% des auditeurs.

Nous avons quand même calculé par curiosité la somme des degrés d'emphase (1, 2 ou 3) pour les mots donnés emphatiques et nous avons remarqué que celle-ci correspond presque parfaitement au pourcentage d'emphase déterminé ci-dessus.

⁴³ Cf. CD ROM pour les feuilles de calcul Excel : *Sondages : Résultats*.

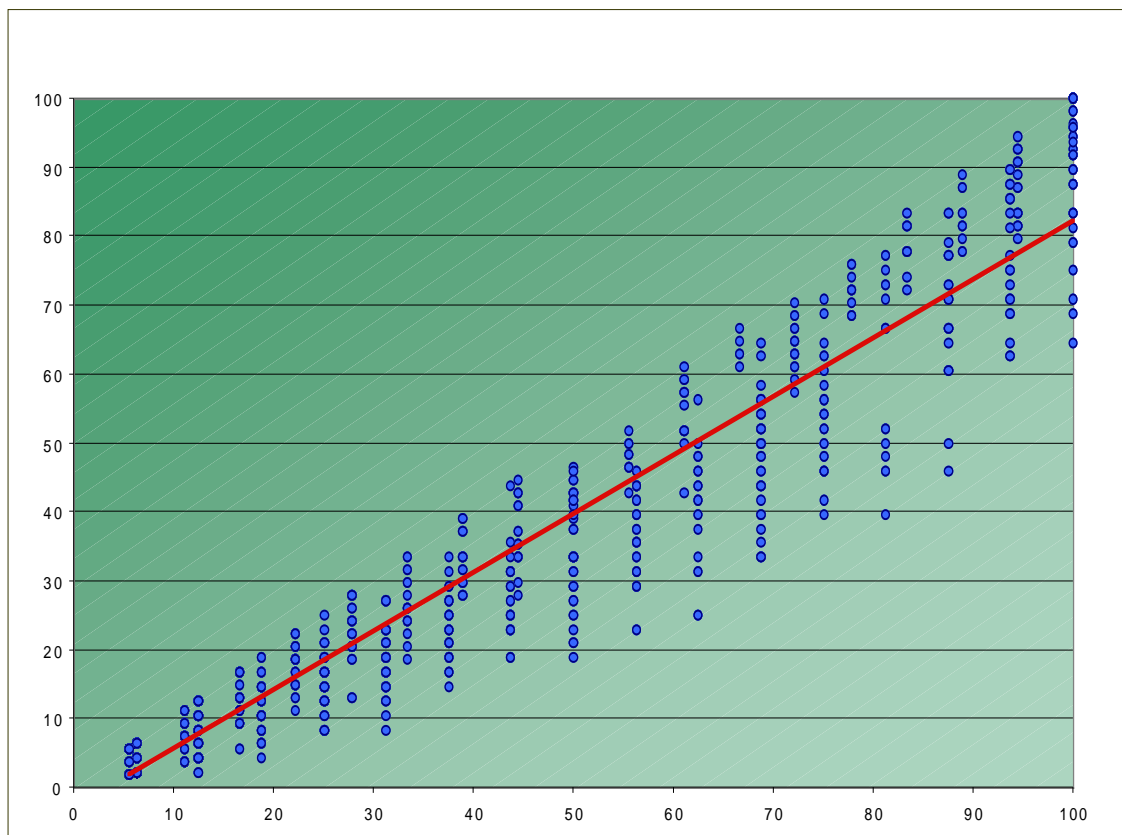
Avec un graphique de type « nuages de points », qui compare des paires de valeurs - ici le pourcentage d'émphase (cf. ci-dessus) et le pourcentage de degré d'émphase (la somme des degrés donnés pour chaque mot, multiplié par 100 et divisé par la somme maximale possible) – nous voyons que les deux valeurs sont très bien corrélées, même pour le sondage 1, dont le graphique suit :



Graphique 1 : corrélation entre le pourcentage d'émphase et le pourcentage de degré d'émphase pour le sondage 1

En ce qui concerne le degré d'émphase, les résultats ne sont donc pas aussi faussés qu'on aurait pu le croire.

Nous avons réalisé le même graphique mais en regroupant cette fois les résultats des trois sondages :



Graphique 2 : Corrélations entre le pourcentage d'émphase et le pourcentage de degré d'émphase pour les sondages 1, 2 et 3 confondus

Comme pour le sondage 1 pris séparément, les nuages de points montrent une assez faible dispersion (un peu plus importante ici quand même) : la ligne rouge est la droite de régression. Ce graphique montre que l'écart par rapport à la droite de régression n'est pas très important et assez constant, ce qui veut dire que quelle que soit la variable que l'on choisit, le résultat sera pratiquement le même. Ceci prouve qu'on arrive à un assez bon consensus pour le jugement de l'émphase. Ce graphique nous permet également de voir qu'il n'existe pas de zone vide, c'est-à-dire que l'émphase ici semble être un phénomène plutôt continu. Nous réservons cependant notre jugement (cf. 2.2.1 et 4.3.5).

Notre idée de départ était de sélectionner, grâce aux sondages, des passages très emphatiques et peu emphatiques et de réaliser un test de perception sur ordinateur pour avoir des résultats plus fiables, tous les auditeurs se trouvant dans les mêmes conditions.

Nous disposons en effet au Laboratoire Parole et Langage de l'Université de Provence d'un logiciel appelé Perceval qui permet de faire des tests de perception : les auditeurs voient une phrase sur l'écran, l'entendent dans le casque et ont la possibilité d'appuyer sur des boutons (cinq boutons seulement) pour appliquer les consignes données. Ceci implique que le test doit être précisément ciblé et qu'il faut poser des questions avec comme réponse oui/non. Il aurait donc fallu faire entendre des phrases et demander à chaque fois si tel mot était emphatique ou non. Nous avons réfléchi à différents types de tests et de questions, au nombre de phrases, etc... et nous avons même préparé ce test mais nous en avons finalement abandonné l'idée. En effet, ce test nous semblait sans intérêt par rapport aux sondages déjà effectués et surtout trop restrictif et directif. Nous aurions alors appliqué les méthodes que nous avons critiquées précédemment.

Les résultats des sondages se sont de plus révélés très pertinents. Plusieurs mots reviennent chez tous les locuteurs. Les statistiques effectuées nous permettent de déterminer un pourcentage d'emphase et un pourcentage de degré d'emphase⁴⁴. Ces résultats nous semblent amplement suffisants car ils sont extrêmement probants. On arrive à des pourcentages très élevés⁴⁵. De plus, les passages sélectionnés nous paraissent représenter des procédés d'emphase très divers (cela sera confirmé ou infirmé par nos analyses).

Pour toutes ces raisons, nous n'avons pas mis en place un test de perception sur ordinateur qui n'aurait rien apporté de plus si ce n'est des restrictions des passages à étudier, ce que nous voulions éviter.

⁴⁴ Cf. ci-dessus.

⁴⁵ Cf. CD ROM : *Sondages : Résultats*.

2.3 *Traitement du signal acoustique*

Avant de commencer l'analyse des paramètres acoustiques, il fallait procéder au traitement du signal. Cela consiste en plusieurs phases :

- tout d'abord la sauvegarde du son sur CD Rom ;
- ensuite l'étiquetage du corpus en passages, en segments puis en mots ;
- enfin la transcription phonétique des passages sélectionnés.

2.3.1 Numérisation

Notre corpus provient d'origines diverses mais nous avons tout sur cassettes audio simples, non numériques. Devant le risque de voir nos données s'abîmer sur des bandes sonores de la sorte et surtout afin de pouvoir ensuite analyser nos données sur des machines informatiques, il nous fallait procéder à la numérisation des données.

La numérisation de nos cassettes audio a été réalisée grâce au logiciel SoundForge, installé sur réseau PC, sous un environnement WINDOWS 95. Un lecteur de cassettes audio est relié à un ordinateur qui transforme le signal analogique en signal numérique. On peut ensuite graver un CD à partir de l'enregistrement numérique, et obtenir un CD ROM. C'est ce que nous avons fait.

2.3.2 Etiquetage du corpus

Dans la mesure où nous avons choisi de fabriquer notre propre corpus et non pas de prendre un corpus déjà existant et donc déjà étiqueté, il nous a fallu réaliser l'étiquetage nous-même.

Nous avons d'abord distingué nos trois corpus (1. le débat politique, 2. la conversation, et 3. l'émission de radio) et numéroté également les passages longs que nous avons fait entendre dans les tests de perception avant de choisir des phrases ou unités intonatives à juger.

Ainsi nous avons donné des noms à nos trois corpus :

- P1. ... signifie qu'il s'agit du débat politique ;
- P2. ... signifie qu'il s'agit de la conversation ;
- P3. ... signifie qu'il s'agit de l'émission radio.

Le chiffre suivant fait référence aux passages longs sélectionnés dans les tests de perception : P1.1 ... ou P3.4 ... ou encore P2.6 ..., etc...

Nous avons ensuite numéroté les segments qui se trouvent dans chaque passage (S01 par exemple) et obtenons donc pour chaque segment un nom comme suit : P1.4S05, par exemple, qui correspond au cinquième segment du passage 4 du débat politique.

2.3.2.1 PRAAT

Afin d'analyser nos données, nous avons choisi d'utiliser le logiciel PRAAT (Boersma & Weenik, 1996). Ce logiciel a été développé par Paul Boersma et David Weenik à l'institut des sciences phonétiques de l'université d'Amsterdam. Il s'agit d'un logiciel extrêmement performant, qui permet d'analyser la parole (spectrogrammes, courbes de fréquence fondamentale, d'intensité, etc...), de la manipuler (possibilité de modifier les contours mélodiques ou les durées par exemple), de segmenter et d'étiqueter de façon tout à fait aisée, en intégrant des symboles phonétiques. PRAAT produit des graphiques de haute qualité et offre la possibilité de programmer des prises de mesure par exemple, ou encore des graphiques, grâce à des scripts. En plus de tous ces atouts, PRAAT est compatible avec n'importe quelle plateforme informatique et permet de naviguer de l'une à l'autre sans aucune difficulté. Il existe aussi la possibilité de faire de la synthèse ainsi que des statistiques, mais nous n'avons pas eu recours à ces deux fonctions. Nous avons opté pour d'autres outils en ce qui concerne la partie synthèse et la partie statistique.

Enfin, en plus de sa performance, PRAAT est le seul logiciel de ce type à être diffusé gratuitement⁴⁶.

⁴⁶ Paul Boersma diffuse PRAAT gracieusement aux chercheurs à condition qu'on lui en fasse la demande en personne : <paul.boersma@hum.uva.nl>. Cf. également <<http://www.fon.hum.uva.nl/praat>> ou <<http://www.praat.org>> pour plus de détails.

2.3.2.2 Passages et segments

Nous avons d'abord découpé nos corpus en passages, puis en segments. Grâce à PRAAT, nous pouvons sélectionner des passages de notre corpus, leur donner un nom et voir surtout leur durée et leur emplacement dans tout le corpus. Ainsi nous avons étiqueté chacun de nos trois extraits de conversation en passages, puis en segments, qui correspondent aux morceaux sélectionnés dans les tests de perception.

Cette manipulation est assez rapide et ne présente pas de difficultés particulières : PRAAT nous permet de visualiser le son et il est possible d'associer au fichier son un fichier nommé « TextGrid », dans lequel on peut étiqueter le son et rentrer du texte. Nous avons effectué cette manipulation pour 19 passages et 146 segments⁴⁷. Voici un exemple (figure 1 ci-dessous). Il s'agit du corpus 1 (le débat politique). C'est notre morceau le plus long. Neuf passages ont été sélectionnés parmi lesquels nous avons choisi 59 segments.

⁴⁷ Cf. CD ROM : *Etiquetage : Son et Etiquetage : TextGrid*.

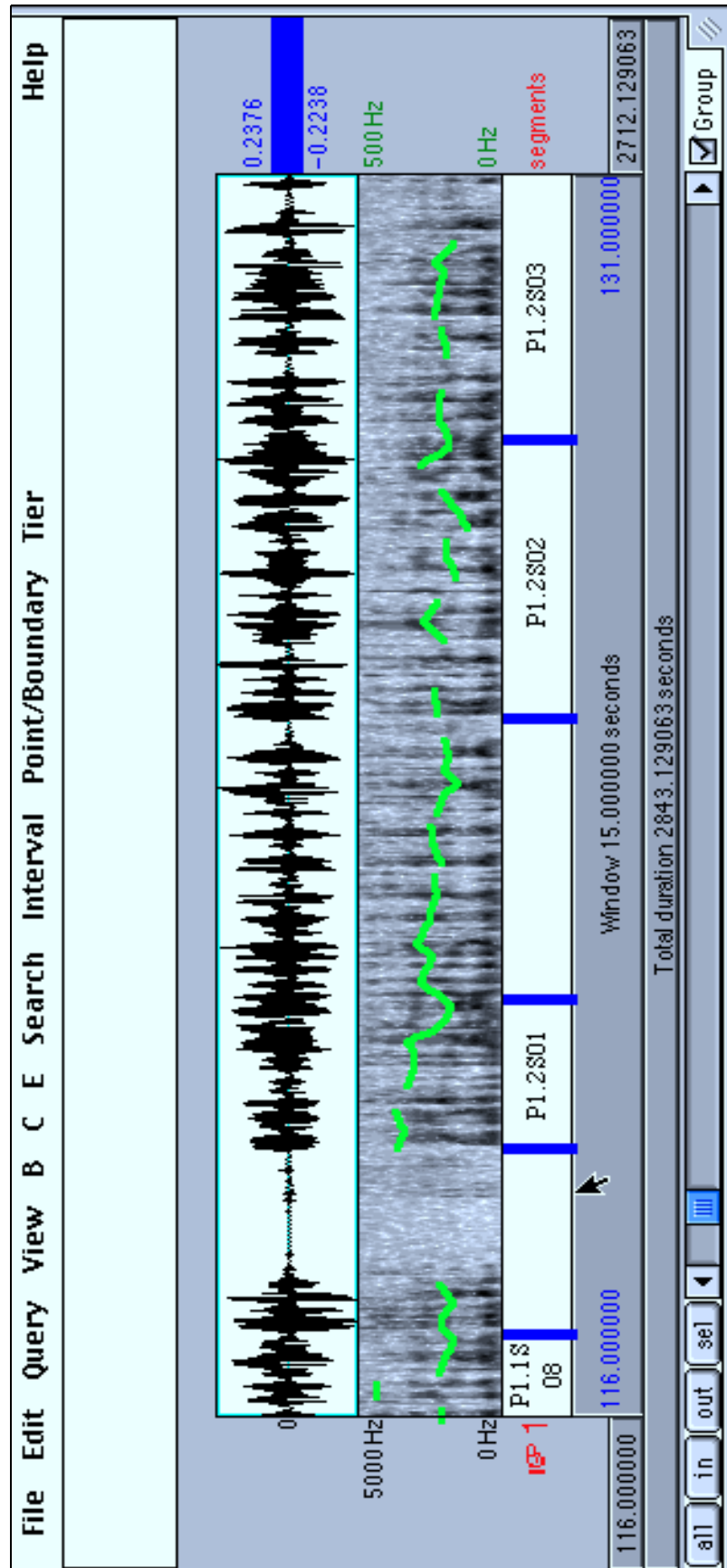


Figure 1 : Question Time

Sur la figure 1, on peut voir l'oscillogramme en haut, puis le spectre en différents niveaux de gris, dans lequel la courbe verte représente la fréquence fondamentale. L'étage qui suit (appelé « tier ») fait partie du fichier « TextGrid » qui est ici couplé au fichier son. Nous avons nommé cet étage « segments ».

2.3.2.3 Mots

La manipulation suivante consistait à étiqueter chaque segment en mots. A nouveau grâce au logiciel PRAAT, nous avons effectué ce travail qui ne peut se faire automatiquement. Nous avons fait cela à l'oreille et à la vue (il est impossible de le faire uniquement à l'oreille, ce serait beaucoup plus long et moins précis) : comme nous venons de le montrer figure 1, PRAAT nous permet en effet de visualiser l'oscillogramme, le spectrogramme (duquel nous enlevions les formants), et la courbe de F0. Mais il est possible aussi d'avoir la courbe d'intensité ainsi que le voisement. C'est grâce à tout cela et à l'écoute, bien entendu, que nous avons pu marquer les frontières des mots sur le signal, en utilisant à nouveau des fichiers « TextGrid », dans lesquels nous avons rentré cette fois les étiquettes des mots et non plus des segments. Nous avons effectué ce travail sur les 146 segments et avons réalisé l'étiquetage d'environ 2400 mots.

Le découpage peut être parfois assez aisé, en particulier s'il s'agit de mots venant juste après une pause. Le début du mot est alors net sur l'oscillogramme (la pause est représentée par un trait continu et dès que le phonème débute, des vibrations sont visibles) ainsi que sur le spectrogramme (qui est alors gris clair puisqu'on n'a pas de son), surtout si le mot commence par une voyelle ou une consonne voisée autre qu'une plosive. Notons que si l'on a du bruit de fond par exemple, il va être représenté sur l'oscillogramme par de très faibles vibrations et sur le spectrogramme par du gris un peu plus foncé mais dans le milieu du spectre, ce qui est un indice de bruit, et non de voyelles (bas du spectre beaucoup plus foncé) ou de sifflante (haut du spectre plus foncé)⁴⁸.

⁴⁸ Afin de mieux analyser le spectre, nous faisons disparaître la courbe de fréquence fondamentale quand cela nous semble approprié.

C'est le cas dans le segment P3.1S06, où la première pause est nette et la deuxième contient un peu de bruit⁴⁹ :

☞ **exs.1** : YB : *Well I think the one thing it achieves, just by (pause) existing (pause) is telling people...*

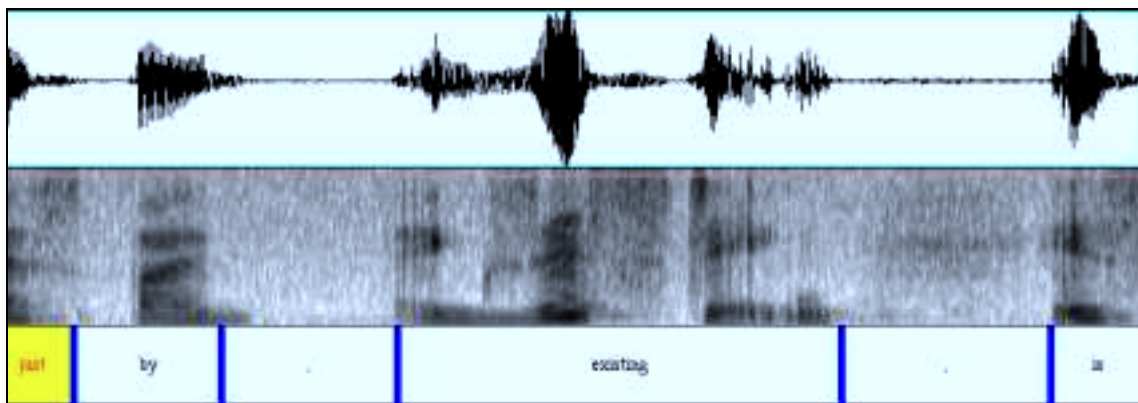


Figure 2⁵⁰ : P3.1S06

Si le mot a une plosive initiale, cela devient alors compliqué car la plosive est caractérisée par une explosion qui est représentée par un blanc sur le sonagramme et rien sur l'oscillogramme, comme dans le segment P3.1S08 :

☞ **exs.2** : YB : *It also liberates I think (pause) policemen ...*

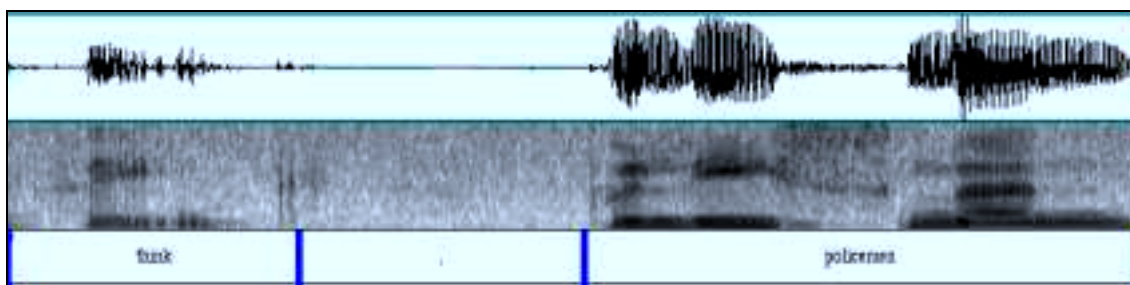


Figure 3 : P3.1S08

⁴⁹ A chaque fois que nous citons un exemple pour lequel nous pensons qu'il est primordial de pouvoir en avoir le son, nous le signalons par le symbole ☞, nous le nommons 'exs.' (exemple sonore) et le numérotions (exs.1 par exemple) afin qu'il puisse être aisé de l'écouter sur le CD ROM. Tous les exemples sonores sont disponibles dans le fichier *exemples sonores* sur le CD ROM. Ils suivent la chronologie du texte. Ce sont des fichiers son .wav.

⁵⁰ Les pauses sont symbolisées ici par une virgule. Comme dans toutes les figures qui vont suivre dans cette sous-section, on peut voir à nouveau l'oscillogramme, le spectrogramme et la courbe de F0 en vert, ainsi que le fichier « TextGrid » collé sous le spectrogramme, qui est composé ici non pas d'un étage nommé « segments » comme dans la figure 1, mais d'un étage nommé « words » dans lequel on a rentré les mots du segment.

Une plosive sera par contre aisément reconnaissable quand elle n'apparaît pas après une pause, dans P2.6S08, par exemple :

☞ **exs.3** : *D* : you start to talk like that **person** talks as well

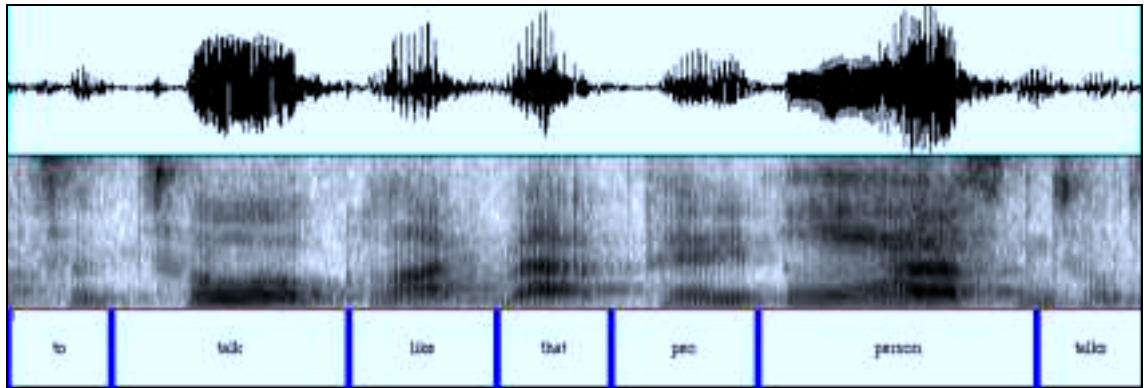


Figure 4 : P2.6S08

Il est également assez facile de trouver le début ou la fin des fricatives qui sont visibles dans le haut du sonagramme, comme le montre la figure suivante pour le segment P2.1S04 :

☞ **exs.4** : *V* : I think **she** did actually suggest that I stay there again this year.

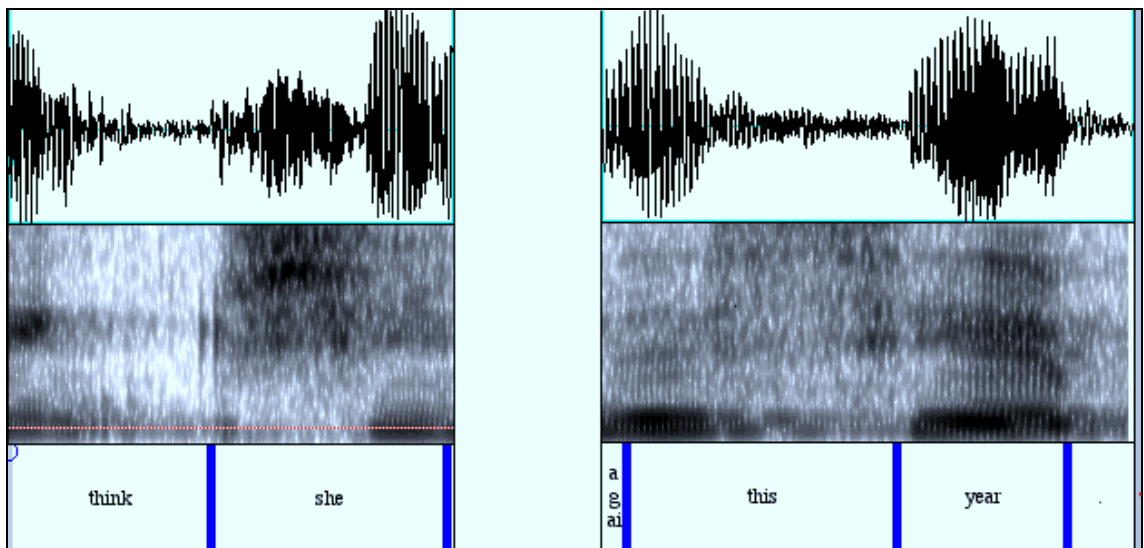


Figure 5 : P2.1S04

Néanmoins il est souvent des cas où le découpage est très difficile, quand le phénomène de gémiation intervient par exemple ou encore quand deux voyelles se suivent, ou quand on a un /r/ suivi d'une voyelle, etc ...

Dans le segment P1.1S01, on trouve même deux diphtongues qui se suivent, ce qui complique encore la tâche :

☞ **exs.5** : DD : *This is a kerfuffle today over whether ...*

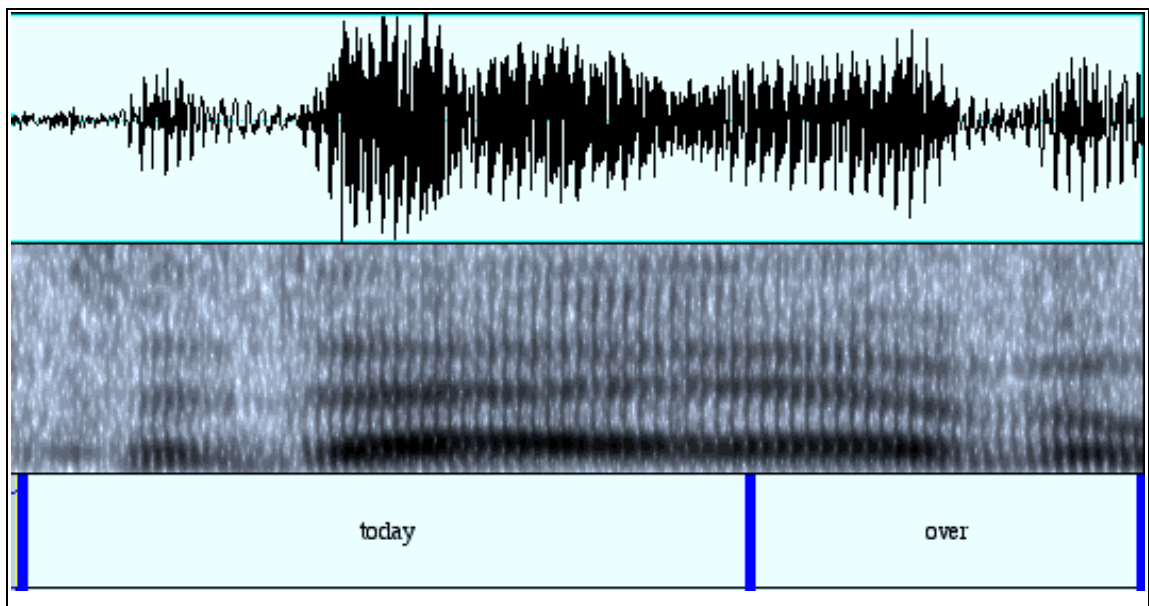


Figure 6 : P1.1S01

Dans le segment P1.5S04, on a une gémiation et aussi /r/ + voy :

☞ **exs.6** :PM : *We have in Ken Clarke the best Chancellor for a generation*

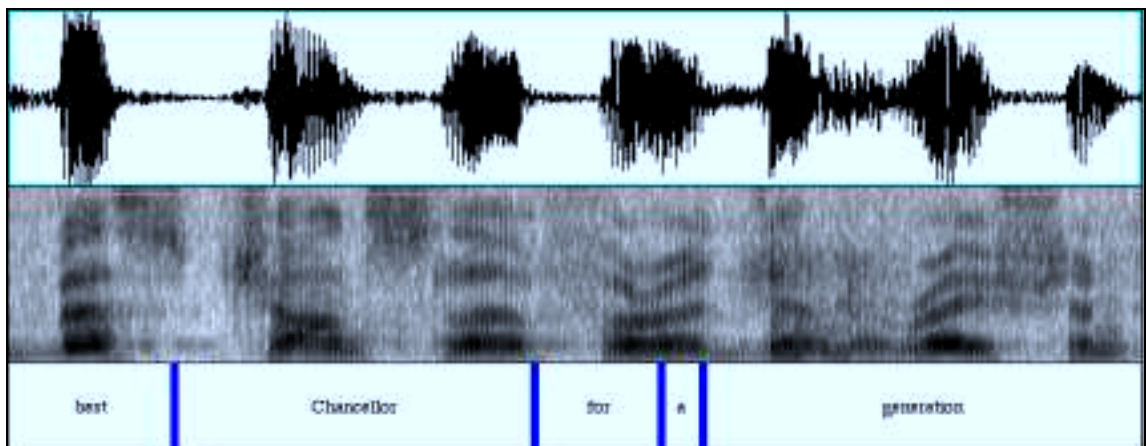


Figure 7 : P1.5S04

Dans ces cas précis, seule l'oreille peut nous permettre de segmenter.

Ce travail d'étiquetage a été très long mais il fallait obligatoirement en passer par là afin de pouvoir ensuite effectuer les mesures acoustiques et aucune autre solution hormis la solution manuelle n'était possible. Au terme de ce travail, nous étions donc en possession de fichiers TextGrid contenant des étiquettes avec des mots pouvant être alignés aux fichiers son correspondants. Les deux types de fichiers sont séparés, mais on peut les voir alignés à l'écran dans une seule fenêtre, comme c'est le cas pour toutes les figures ci-dessus.

Voici les fichiers séparés :

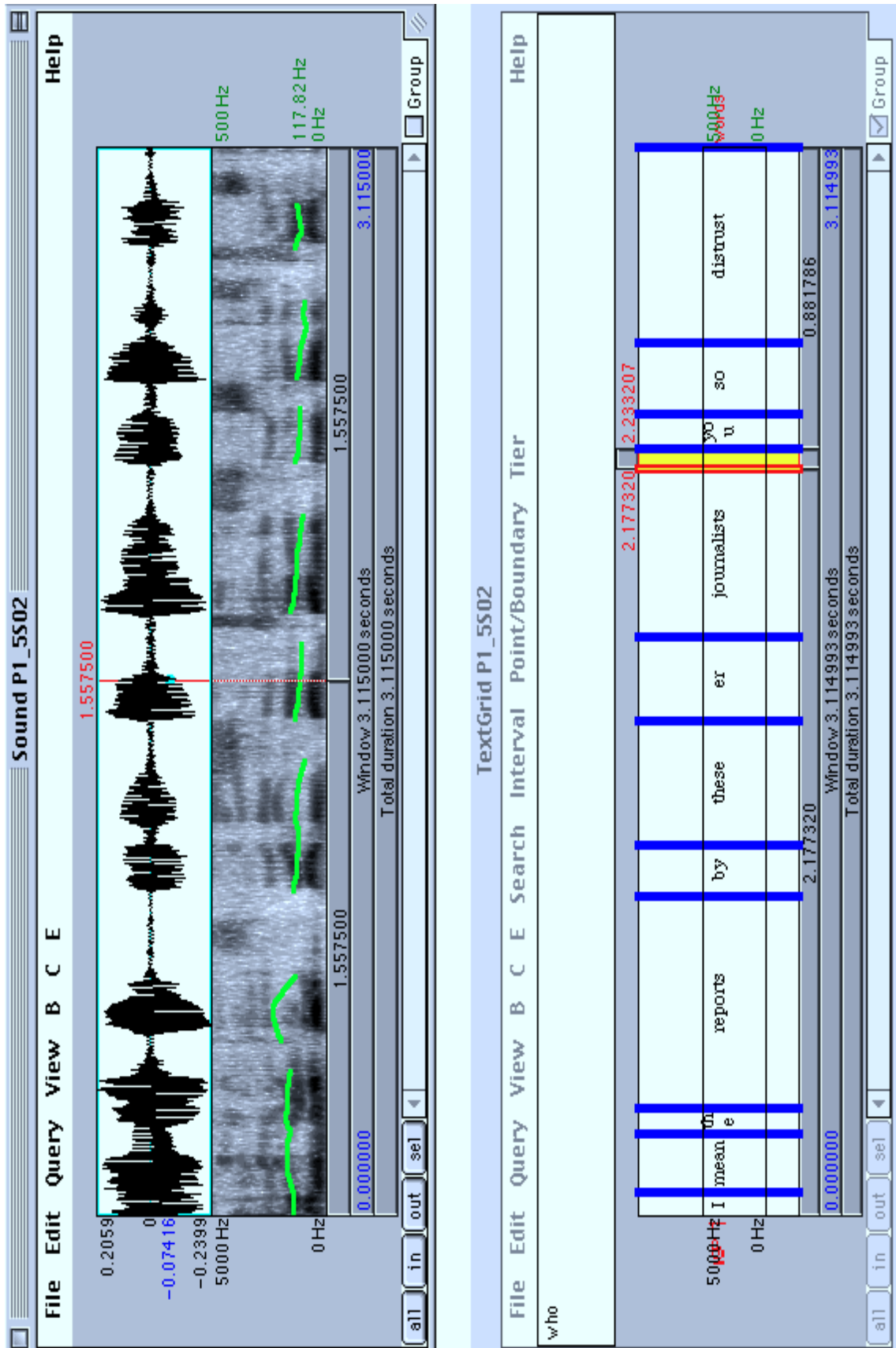


Figure 8 : P1.5S02 : fichier son et fichier TextGrid

2.3.2.4 « Prominence »

Une fois en possession de nos fichiers TextGrid contenant les étiquettes pour les mots, nous avons ajouté à ces mêmes fichiers un étage (« tier ») sous celui des mots, que nous avons nommé « prominence », et dans lequel nous avons rentré le pourcentage d'emphasis et le pourcentage de degré d'emphasis qui correspondait à chaque mot.

Nous avons rentré les deux pourcentages dans le même étage en les séparant simplement par une barre, comme on peut le voir dans l'exemple ci-dessous⁵¹ :

⁵¹ Tous les fichiers sont disponibles sur le CD ROM : *Etiquetage : Son* pour le son (.wav) et *Etiquetage : Textgrid* pour les fichiers comprenant le découpage en mots et l'étage avec le degré d'emphasis (« prominence tier ») en plus de celui des mots. Ces fichiers sont des fichiers PRAAT. Nous n'avons volontairement pas mis PRAAT à disposition sur le CD ROM, par respect pour Paul Boersma qui souhaite qu'on le lui demande en personne (cf. note 46). Cependant, ces fichiers peuvent être ouverts grâce à d'autres logiciels, comme *BEdit* pour les TextGrid ou *QuickTime* pour le son par exemple. *BEdit* est disponible sur le CD ROM.

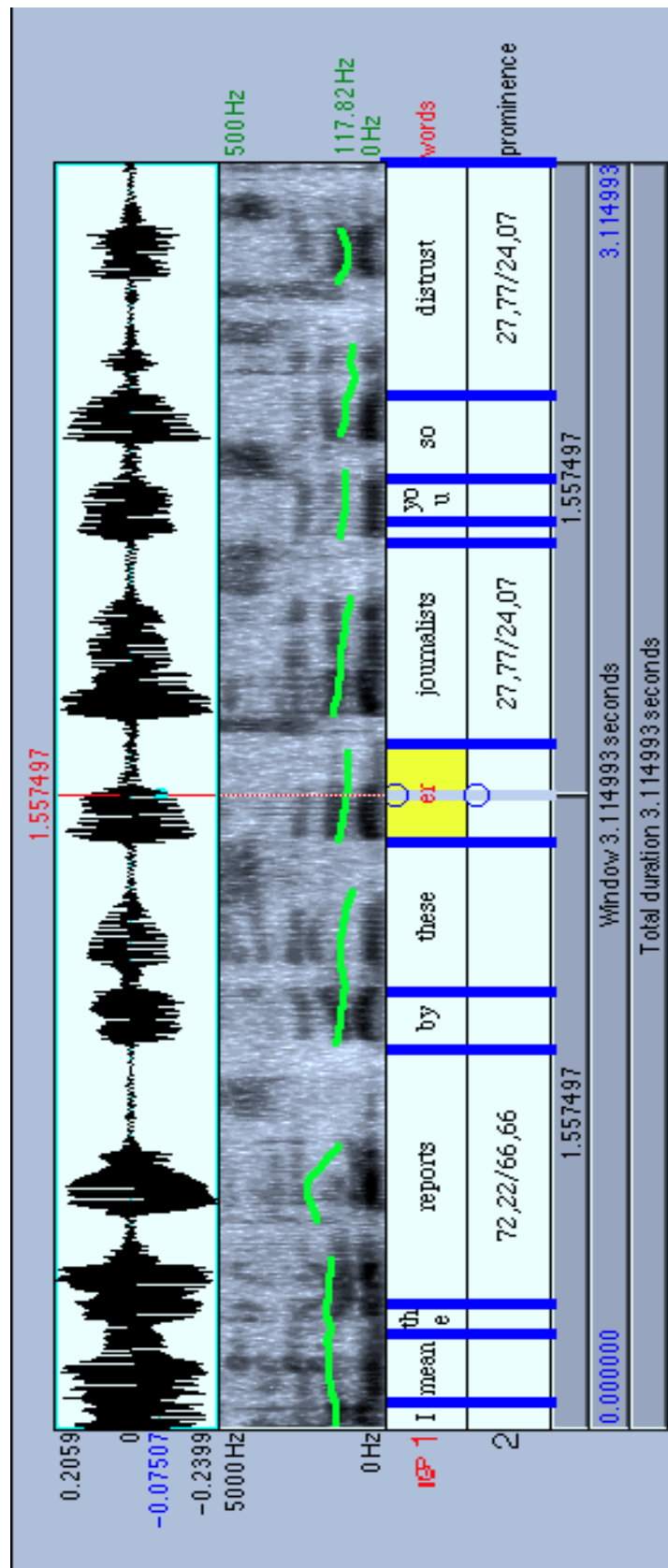


Figure 9 : P1.5S02

2.3.3 Transcription phonétique

En décembre 2000, Nick Campbell⁵² est venu passer un mois au laboratoire Parole et Langage d'Aix-en-Provence et y a implémenté un dictionnaire de phonétique automatique. A partir de l'orthographe du mot, la transcription phonétique est donnée automatiquement. Ce dictionnaire comportait des lacunes que Bouzon (2001) et Auran (doctorat en cours) se sont évertués à combler. Ils ont rajouté des mots au dictionnaire, modifié les symboles phonétiques en symboles SAMPA⁵³ (Wells *et al.*, 1992 ; Wells, 1995 ; SAMPA, [http : cf. références bibliographiques](http://cf.reférences.bibliographiques)) et surtout complété ce dictionnaire en rajoutant au moyen de scripts⁵⁴ les formes neutralisées des mots outils (« of » et « for » principalement), la transcription des chiffres et des nombres, les différentes formes des articles « the » et « a » (devant voyelle, devant consonne ou mis en relief), ou encore le « r » de liaison (cf. Bouzon, 2001 pour tous les détails).

Nous avons bénéficié de cet outil tout récent. Les mots que nous avons rentrés dans nos fichiers PRAAT pour l'étiquetage en mots ont donc été transcrits en phonétique automatiquement.

Mais avant de pouvoir effectuer cette manipulation, nous devons procéder à la mise en forme de nos fichiers.

2.3.3.1 Mise en forme des fichiers

Nous avons deux tâches à réaliser :

tout d'abord reprendre tous les mots entrés dans nos fichiers TextGrid PRAAT et noter ceux dont la prononciation était réduite. Nous avons donc profité du deuxième étage où nous avons noté les pourcentages d'emphase, et nous avons mis un point dans cet étage-là (nommé « prominence ») sous tous les mots réduits, les mots donnés emphatiques n'étant logiquement pas réduits mais prononcés dans leur forme pleine. C'est pourquoi nous pouvons normalement utiliser le même

⁵² Nick Campbell est chercheur au Japon pour ATR ; <<http://www.isd.atr.co.jp/esp>>.

⁵³ Caroline Bouzon et Cyril Auran ont choisi la notation phonétique SAMPA car elle est compatible avec n'importe quel ordinateur et n'importe quel clavier et devient la notation phonétique universelle. La correspondance entre les symboles de l'alphabet phonétique international (IPA) et l'alphabet SAMPA est donnée en annexe 1 ainsi que sur le CD ROM : *IPA/SAMPA*.

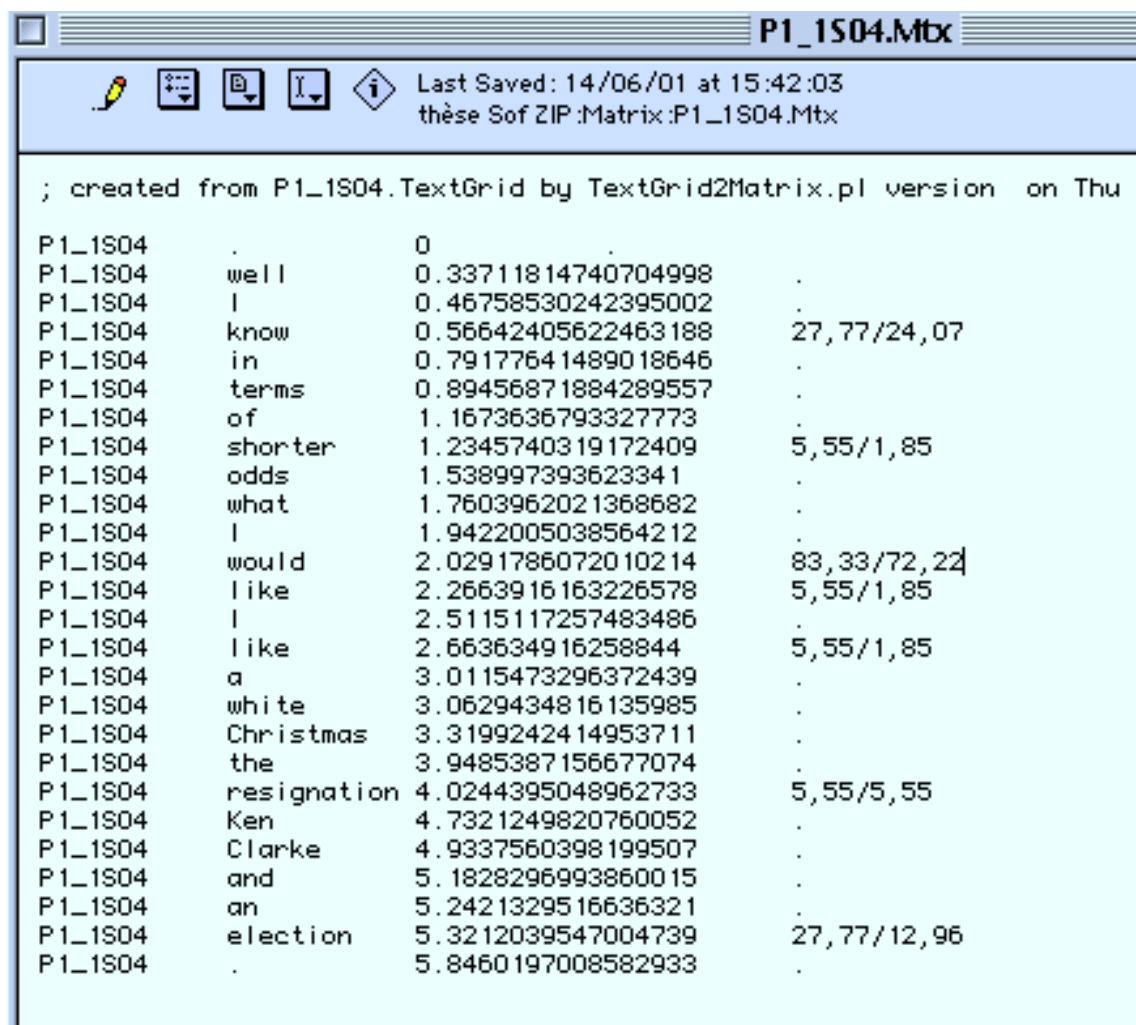
⁵⁴ Les scripts ont été rédigés en langage PERL.

étage. Or nous nous sommes trouvée face à un problème : certains mots, des articles et des occurrences de « of » en particulier, sont prononcés de façon réduite et sont pourtant donnés emphatiques par une ou deux personnes. Leur forme phonétique étant réduite, nous devions noter un point pour que la transcription phonétique soit correcte, mais nous devions également noter un pourcentage d'emphase. Nous avons regardé ces mots de plus près et avons constaté qu'il s'agissait toujours de mots entourant des mots emphatiques à un fort degré. L'article d'un mot emphatique par exemple est prononcé sous sa forme réduite et est quand même perçu emphatique par quelques auditeurs. Cette remarque pourra nous être utile par la suite, mais à ce stade, il fallait prendre une décision : nous avons décidé que lorsqu'un mot était prononcé sous sa forme réduite, nous ne marquerions pas de pourcentage d'emphase sauf si celui-ci dépassait deux locuteurs (sur 18 dans le premier sondage et sur 16 dans les deux autres), auquel cas la transcription phonétique serait erronée, mais nous pourrions la reprendre manuellement sans avoir omis le degré d'emphase.

La deuxième tâche qui nous incombait était de fournir des fichiers possédant un format compatible avec le dictionnaire automatique. Nos fichiers TextGrid PRAAT ne convenaient pas. Daniel Hirst a réalisé pour nous un script en langage Perl, *TextGrid2Matrix*⁵⁵, qui convertit nos fichiers TextGrid en fichiers Matrix (Mtx) compatibles avec le dictionnaire et dont la figure 10 montre un exemple⁵⁶. La colonne de gauche correspond au nom du segment, la colonne suivante contient les mots, puis on a le début en secondes de chaque mot, puis le contenu de l'étage nommé « prominence » dans les fichiers TextGrid, à savoir le pourcentage d'emphase et de degré d'emphase. Quand aucun pourcentage n'est donné, on a un point, ce qui signifie que si le mot possède une forme phonétique réduite, c'est celle-ci qui sera choisie lors de l'étape suivante.

⁵⁵ Cf. CD ROM : *Scripts : PERL : TextGrid2Matrix* ou annexe 2c.

⁵⁶ Cf. CD ROM : *Etiquetage : Mtx* pour tous les autres fichiers .Mtx.



```

; created from P1_1S04.TextGrid by TextGrid2Matrix.pl version on Thu .
P1_1S04 . 0
P1_1S04 well 0.33711814740704998 .
P1_1S04 I 0.46758530242395002 .
P1_1S04 know 0.56642405622463188 27,77/24,07
P1_1S04 in 0.79177641489018646 .
P1_1S04 terms 0.89456871884289557 .
P1_1S04 of 1.1673636793327773 .
P1_1S04 shorter 1.2345740319172409 5,55/1,85
P1_1S04 odds 1.538997393623341 .
P1_1S04 what 1.7603962021368682 .
P1_1S04 I 1.9422005038564212 .
P1_1S04 would 2.0291786072010214 83,33/72,22
P1_1S04 like 2.2663916163226578 5,55/1,85
P1_1S04 I 2.5115117257483486 .
P1_1S04 like 2.663634916258844 5,55/1,85
P1_1S04 a 3.0115473296372439 .
P1_1S04 white 3.0629434816135985 .
P1_1S04 Christmas 3.3199242414953711 .
P1_1S04 the 3.9485387156677074 .
P1_1S04 resignation 4.0244395048962733 5,55/5,55
P1_1S04 Ken 4.7321249820760052 .
P1_1S04 Clarke 4.9337560398199507 .
P1_1S04 and 5.1828296993860015 .
P1_1S04 an 5.2421329516636321 .
P1_1S04 election 5.3212039547004739 27,77/12,96
P1_1S04 . 5.8460197008582933 .

```

Figure 10 : fichier .mtx P1.1S04

Ces deux tâches effectuées, un nouveau script Perl (*AddPron2Matrix*⁵⁷), également écrit par Daniel Hirst, nous a permis de rajouter la phonétique à nos fichiers Mtx, en les convertissant en fichiers .Mtx « enrichis », que nous appelons ‘Mtx + phon’⁵⁸. La figure 11 montre le même fichier que dans la figure 10, enrichi de la phonétique en alphabet SAMPA dans la cinquième colonne, des références des différents dictionnaires qui ont permis de trouver la phonétique dans la colonne qui suit. La dernière colonne, composée de chiffres, correspond au nombre de syllabes du mot.

⁵⁷ Cf. CD ROM : *Scripts : PERL : AddPron2Matrix* ou annexe 2d.

⁵⁸ Cf. CD ROM : *Etiquetage : Mtx+phon* pour tous les autres fichiers .Mtx ‘enrichis’ de la phonétique.

P1_1S04.Mtx		Last Saved: 6/07/01 at 15:15:04		P1_1S04.Mtx	
Macintosh HD:Thèse Sof :Mtx+phon:P1_1S04.Mtx					
; pronunciation etc added to P1_1S04.Mtx by S:\EnglishProsodyGroup\Marsek\scripts\perl\AddPron2Matrix.pl version					
; created from P1_1S04.TextGrid by TextGrid2Matrix.pl version on Thu Jun 14 15:42:02 2001					
P1_1S04	.	0	.	-	PNCT
P1_1S04	well	0.33711814740704998	.	wel	l0%,K6%,0p%,Pu*,M-%
P1_1S04	I	0.46758530242395002	.	ai	Ki\$,Qx* 1
P1_1S04	know	0.56642405622463188	27,77/24,07	n0U	J5*,Ki% 12R,3R,6R,8,9,10
P1_1S04	in	0.79177641489018646	.	ln	K6%,P+*,T-* 1
P1_1S04	terms	0.89456871884289557	.	t3:mz	Ha%,Kj% 123
P1_1S04	of	1.1673636793327773	.	0f	RF 1
P1_1S04	shorter	1.2345740319172409	5,55/1,85	's0:t0r*	Or% 2
P1_1S04	odds	1.538997393623341	.	Qdz	Kj% 1
P1_1S04	what	1.7603962021368682	.	wQt	0R*,Qy*,Qz* 1
P1_1S04	I	1.9422005038564212	.	ai	Ki\$,Qx* 1
P1_1S04	would	2.0291786072010214	83,33/72,22	wUd	Gc*,Hc% 15,6R,9
P1_1S04	like	2.2663916163226578	5,55/1,85	laik	H2*,K6%,0R%,Pu%,T-*,V-* 1E
P1_1S04	I	2.5115117257483486	.	ai	Ki\$,Qx* 1
P1_1S04	like	2.663634916258844	5,55/1,85	laik	H2*,K6%,0R%,Pu%,T-*,V-* 1E
P1_1S04	a	3.0115473296372439	.	0	RF 1
P1_1S04	white	3.0629434816135985	.	wait	M6%,0B* 1
P1_1S04	Christmas	3.3199242414953711	.	'krism0s	K7%,No% 2
P1_1S04	the	3.9485387156677074	.	Dl*	RF 1
P1_1S04	resignation	4.0244395048962733	5,55/5,55	,rezig'neiSn	M6% 4
P1_1S04	Ken	4.7321249820760052	.	ken	Nl% 1
P1_1S04	Clarke	4.9337560398199507	.	klA:k	** 0
P1_1S04	and	5.1828296993860015	.	0n	RF 1
P1_1S04	an	5.2421329516636321	.	0n	RF 1
P1_1S04	election	5.3212039547004739	27,77/12,96	l'lekSn	M6% 3
P1_1S04	.	5.8460197008582933	.	-	PNCT

Figure 11 : fichier .mtx 'enrichi' (mtx+phon) P1.1S04

Nous avons donc réussi à obtenir la transcription phonétique automatique de nos fichiers. Bien sûr nous avons dû reprendre cette transcription pour corriger les éventuelles erreurs et faire les modifications nécessaires.

Corrections manuelles⁵⁹

Certains mots de notre corpus ne figuraient pas dans le dictionnaire, les noms propres en particulier (« Yasmin » ou « Clarke ») et quelques mots peu courants ou expressions familières (« kerfuffle », « collude », « knee-jerk »), que nous avons rajoutés au dictionnaire afin de le compléter et d'en faire bénéficier les prochains utilisateurs.

Les hésitations dans le discours spontané sont courantes, on fait des faux départs, on se reprend, on ne finit pas un mot. Toutes ces petites choses ne figurent pas à l'évidence dans le dictionnaire et nous avons donc repris manuellement tous les segments. Dans P1.1S03 par exemple (☞ **exs.7**), le locuteur bute un peu sur les mots :

... t to get his k get the scooters his kids' scooters...

Le « t » et le « k » sont transcrits /ti:/ et /keI/ forcément, nous avons donc changé cela.

Nous avons déjà mentionné le problème des mots possédant des formes réduites, prononcés dans leur forme réduite et qui ont pourtant un degré d'emphase. Dans ce cas, nous avons gardé l'emphase, le mot n'apparaît donc pas sous sa forme réduite. Nous avons changé cela aussi :

☞ **exs.8** : P1.7S01 : ...himself to be

« be » est donné emphatique à 94,44/81,48% et « to » à 55,55/55,55% alors qu'il est prononcé /t@/ et non /tu:/. De même dans P2.5S05

☞ **exs.9** : she gave me my PhD

« she » est emphatique à 31,25/27,08% et est en effet prononcé /Si:/, alors que « me » est donné emphatique à 12,5/8,33% et est réduit (/mI/ et non /mi:/).

Le dictionnaire n'intègre pas de composante syntaxique ou grammaticale, ni sémantique, et cela contribue à créer des erreurs.

Pour « have », une seule prononciation est donnée, celle du verbe. Il se trouve qu'il est souvent réduit lorsqu'il est auxiliaire. Dans

⁵⁹ Dans cette sous-section, comme dans tout le reste de notre travail, la phonétique est donnée en alphabet

☞ **10** : P1.8S04 : ...from the one that this lot have had

« have » est donné /h{v/ alors qu'il est en fait prononcé /@f/. Il en va de même dans

☞ **exs.11** : P3.3S02 : ...what would you have done

où « have » est prononcé /@v/ et non /h{v/. Nous avons aussi rencontré le cas inverse, c'est-à-dire qu'un mot qui a une forme réduite n'est pas donné emphatique, et est donc considéré réduit et a pourtant dans le dialogue sa forme pleine :

☞ **exs.12** : P3.3S10 : All families are good ...

« are » n'est pas donné emphatique et est pourtant prononcé /A:/, mais donné /@/ automatiquement. Nous trouvons la même chose en P3.3S06 (☞ **exs.13**) pour le mot « for » qui n'est pas donné emphatique mais est prononcé /fO:/, ceci dû à une hésitation. Enfin un dernier cas intéressant est à noter en ce qui concerne les réductions : nous avons une occurrence de « don't know » prononcé /d@n@U/ (P2.5S10, ☞ **exs.14**), or nous avons deux étiquettes, une pour « don't » et une pour « know », ce qui donne pour le dictionnaire automatique : /d@Untn@U/.

Certains mots se prononcent différemment selon qu'ils sont nom ou verbe et nous en avons deux cas dans le corpus primaire. Le mot « abuse » est employé comme nom dans nos exemples (P3.1S03 et P3.2S02), et transcrit automatiquement comme s'il s'agissait du verbe /@'bju:z/ au lieu du nom /@'bju:s/. Le mot « subject » (P3.3S06) est aussi transcrit comme le verbe /s@'bdZekt/ alors qu'il s'agit du nom /'sVbdZekt/ ici.

On note des erreurs dans le cas d'homographes comme « put » : /pUt/ ou /pVt/. Etrangement selon nous, la deuxième prononciation est donnée, c'est pourtant la moins courante. Pour « live », on trouve /laIv/, qui n'allait pas dans notre segment (P2.6S06) puisqu'il s'agissait du verbe à l'infinitif /IIv/. Enfin le mot « lead » est le plomb /led/ dans P1.5S06 et est transcrit automatiquement comme le verbe /li:d/.

Une autre source d'erreur vient du fait que des variantes existent quant à la prononciation de certains mots. C'est le cas de « year », donné /j3:/ par le dictionnaire alors que nos locuteurs prononcent presque tous /jI@/. De même, le dictionnaire donne /kri:'eIt/ pour « create » et la locutrice dit /krI'eIt/ (P3.3S05). Pour « suicide » (P3.1S03), le dictionnaire donne /'sju:IsaId/, YB ne prononce pas le /j/. De même VB

prononce « studio » (P2.1S03) sans /j/ alors que la prononciation donnée automatiquement est /'stju:dl@U/.

Les assimilations et les effets de coarticulation ne sont pas pris en compte bien évidemment, nous en avons quelques cas ici : DL palatalise le /s/ de « stressing » et prononce /StresIN/ (P2.6S09, ¶ **exs.15**) et VB succombe à la coarticulation classique et extrêmement courante dans l'expression « this year » : /DISjI@/ (P2.1S04, ¶ **exs.16**).

Enfin nous avons rencontré trois cas de glissement accentuel, sur « unemployment » (P1.5S06, ¶ **exs.17**), l'accent lexical se déplace sur la première syllabe et sur « fourteen » et « fifteen » (P3.3S08, ¶ **exs.18**), l'accent remonte à la première syllabe également.

Nous avons effectué toutes les modifications nécessaires et avons ainsi obtenu la transcription phonétique de notre corpus primaire.

2.4 *Résumé et conclusion*

Nous avons choisi de travailler sur une base de données de parole spontanée uniquement. La méthode choisie afin de juger de l'emphase est celle qui consiste à demander l'avis d'anglophones natifs (« native speakers »), et non pas de partir de préconçus ou de jugements personnels.

Les sondages de départ se sont en fait avérés probants et nous n'avons pas ressenti le besoin de mettre d'autres tests en place afin de déterminer quels passages ou mots étaient emphatiques et lesquels ne l'étaient pas dans notre corpus.

Ainsi, notre corpus primaire est constitué des phrases ou unités intonatives que nous avons demandé à nos auditeurs de juger emphatiques ou non, et la façon dont nous avons construit les tests constitue le point de départ de nos étiquetages en passages et en segments.

Les résultats des sondages sont, quant à eux, le point de départ de toutes nos analyses de corrélation. Nous partirons en effet du pourcentage d'emphase⁶⁰ déterminé grâce aux auditeurs pour savoir quels paramètres acoustiques et prosodiques y sont les mieux corrélés.

Ces tests de perception sont donc essentiels puisqu'ils constituent le point de départ indispensable à notre étude des paramètres acoustiques et prosodiques de l'emphase.

La méthode choisie nous semble être la plus objective possible. N'étant pas nous-même anglophone, et l'emphase semblant, d'après les études décrites dans le chapitre 1 et aussi d'après les résultats de nos sondages, être une notion tout à fait subjective et variant d'un auditeur à l'autre, nous pensons que le fait d'avoir l'avis de « native speakers » uniquement est la meilleure solution.

Nous baserons donc toutes nos études sur ces résultats.

⁶⁰ et non du pourcentage de degré d'emphase. Nous avons fait ce choix car nous avons vu (2.2.3.3, graphique 2) que les deux pourcentages sont bien corrélés. Il semblait donc redondant de procéder à deux analyses différentes. De plus, le fait de déterminer un pourcentage d'emphase en revient à déterminer un degré d'emphase. En effet, nous considérons que les mots donnés par 100% des auditeurs sont emphatiques à un haut degré, alors que ceux donnés par 25% des auditeurs par exemple, ne sont que très peu emphatiques (cf. 2.2.3.3). Nous n'éliminons cependant pas le pourcentage de degré d'emphase, qui pourra parfois être indicatif.

Le corpus a été étiqueté en mots. Pour chaque mot, le degré d'emphasis est donné et nous avons également la transcription phonétique. Le corpus est désormais prêt à être analysé. La suite de ce travail décrit tout le processus d'analyse.

3 Chapitre 3

Paramètres acoustiques et prosodiques

Les deux premiers chapitres de notre travail ont permis de déterminer quels sont les paramètres les plus exploitables, et nous avons décrit tout le travail de préparation du corpus ainsi que la méthodologie de départ.


Ce troisième chapitre s'intéresse plus précisément aux paramètres qui semblent pertinents en rapport avec l'emphase. Il se divise dès lors en quatre sous-sections qui traitent des quatre paramètres choisis :

- la fréquence fondamentale ;
- la durée ;
- l'intensité ;
- les pauses.

Le but de notre travail est de déterminer les paramètres les mieux corrélés à l'emphase. Nous avons opté pour une analyse statistique que nous décrirons en détails dans le chapitre 4. Mais afin de pouvoir utiliser l'outil statistique, un processus de préparation est nécessaire. Il s'agit de rentrer des données dans un tableau, et c'est là que se situe tout le problème : quelles mesures faut-il prendre ? Comment effectuer ces mesures ? Quelle unité rythmique de base devons-nous choisir ? Comment mesurer l'intensité ? Comment distinguer les différents types de pause ? etc...

C'est tout ce processus de réflexion et d'analyses, de prises de mesures et de prises de position que ce chapitre décrit. Nous expliquons comment nous effectuons nos mesures, sur quelles bases (et ceci peut varier en fonction des paramètres), quels calculs nous choisissons, et nous tentons à chaque fois de justifier nos choix.

Dans une cinquième sous-section, nous traitons de cas particuliers dont les caractéristiques pertinentes ne pourront être intégrées à l'analyse statistique.

Comme c'est le cas dans le chapitre précédent, lorsque nous citons un exemple qu'il nous semble nécessaire d'entendre, nous le signalons par le symbole  , le nommons

‘**exs.**’ (exemple sonore) et le numérotions (**exs.22** par exemple) afin qu’il puisse être trouvé facilement sur le CD ROM joint (dans le fichier *exemples sonores*), et écouté.

3.1 *Fréquence fondamentale*

La fréquence fondamentale est un paramètre incontournable. Il est reconnu par tous comme primordial dans la perception de la proéminence et de l'emphase.

Nous avons vu que le pic intonatif était essentiel mais qu'il ne fallait pas négliger pour autant tout le contour intonatif de l'unité car il pouvait se montrer très révélateur. Nous avons également parlé de variabilité entre les locuteurs, il nous faudra prendre toutes ces notions en compte. Avant de décider quels paramètres insérer dans notre tableau de données, et même si les études antérieures nous donnent des indices, nous avons fait une étude préliminaire afin d'avoir une idée des paramètres pertinents dans l'analyse de la F0.

3.1.1 Comment analyser la fréquence fondamentale ?

3.1.1.1 F0 brute

Le logiciel PRAAT (Boersma & Weenik, 1996) donne la courbe de F0 brute des segments choisis.

Mais les courbes de fréquence fondamentale ne sont pas le reflet exact des variations de hauteur mélodique que nous percevons. Il est donc difficile d'effectuer des mesures sur ce type de courbe :

the main problem is that these measurements typically show a great variety of irregularities, amongst which it is almost impossible to perceive any speech melody. ('t Hart, 1984 : 194)

Gimson (1987 : 27) nous met en garde contre les sons non voisés qui peuvent « fausser » les courbes :

Our impression of pitch changes in speech results largely from fundamental frequencies carried by vowels and other voiced sounds, the voiceless sounds being discounted. With pitch, as with quality, an instrumental analysis does

not make this human linguistic discrimination and may, therefore, provide us with a good deal of irrelevant information.

Nous avons cependant regardé la F0 brute afin de nous rendre compte par nous-même de ces variations. Nous avons remarqué que d'autres phénomènes faussaient les courbes :

- certains /r/ sont anormalement placés dans l'aigu comme dans P1.7S04, par exemple :

☞ **exs.19** : PM : *I'm not gonna even answer it*

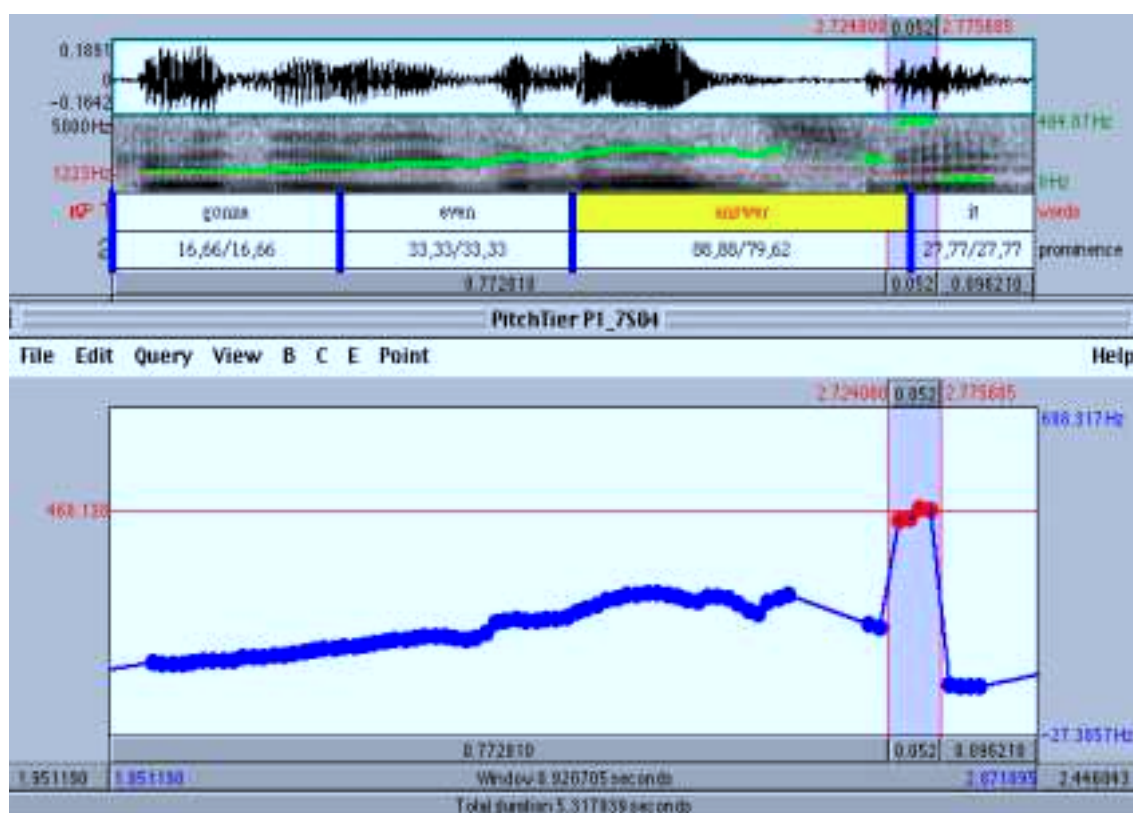


Figure 12⁶¹ : P1.7S04, F0 brute

⁶¹ Cette figure représente ce que PRAAT permet de visualiser. La première fenêtre est celle décrite pour les figures 1 et 2 du chapitre 2, avec l'oscillogramme en haut, le spectre en dessous (les différents tons de gris) avec la courbe de fréquence fondamentale en vert au milieu du spectre, l'étage avec les mots et l'étage avec les pourcentages d'emphase et de degré d'emphase. La deuxième fenêtre (« pitchtier ») montre la courbe brute de fréquence fondamentale toute seule (il s'agit de la même que celle en vert au-dessus, mais grossie, et les points sont reliés par des droites).

- les coups de glotte apparaissent aussi très haut dans les courbes, dans P1.1S08 (☞ **exs.20**) : MM : *a pa / ?ʔ chance* ; dans P1.2S03 (☞ **exs.21**) : MM : *rabbit / ?ʔ* ; ou encore dans P2.4S03 (☞ **exs.22**) : J : *flight / ?ʔ*

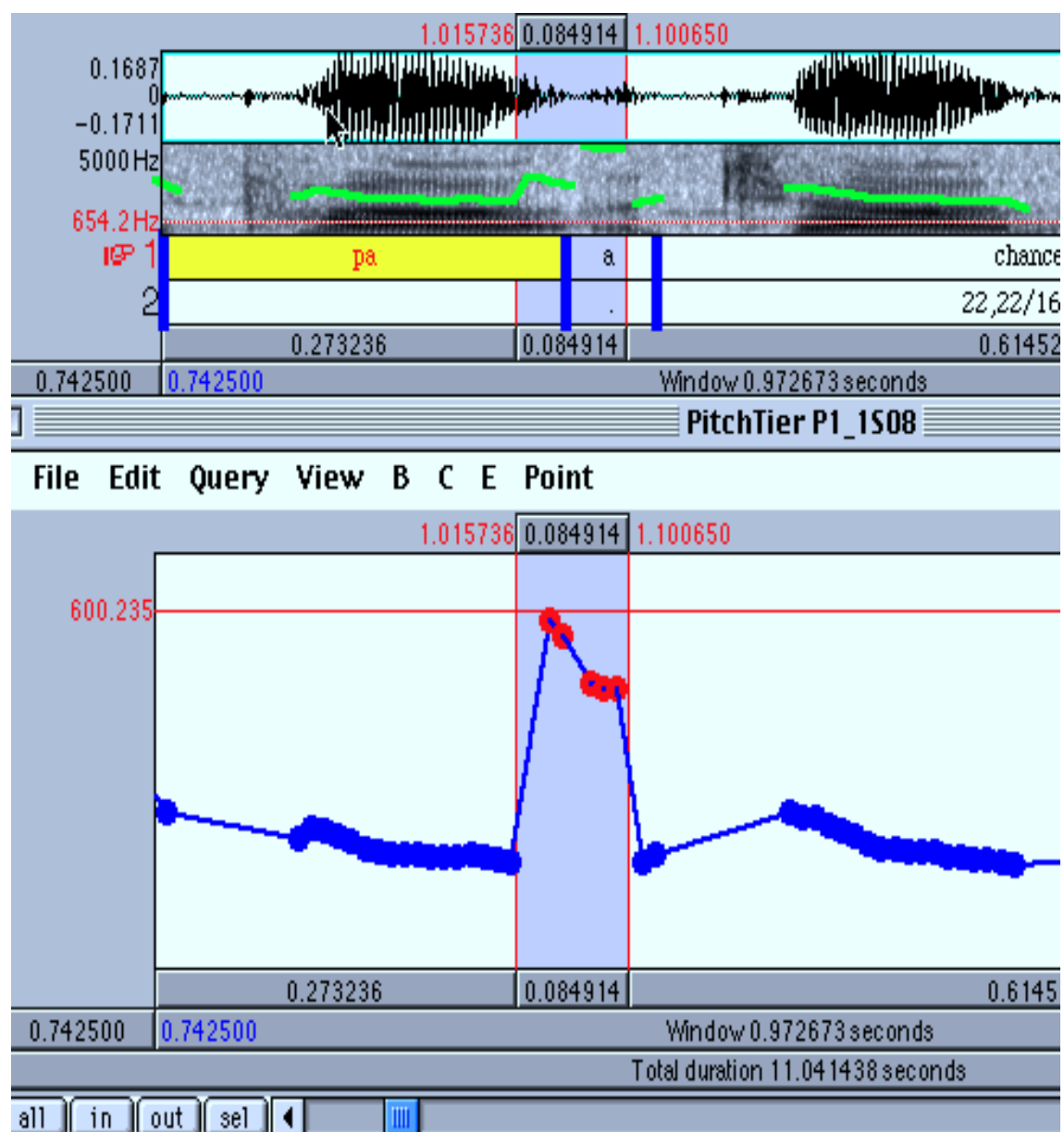


Figure 13 : P1.1S08, F0 brute

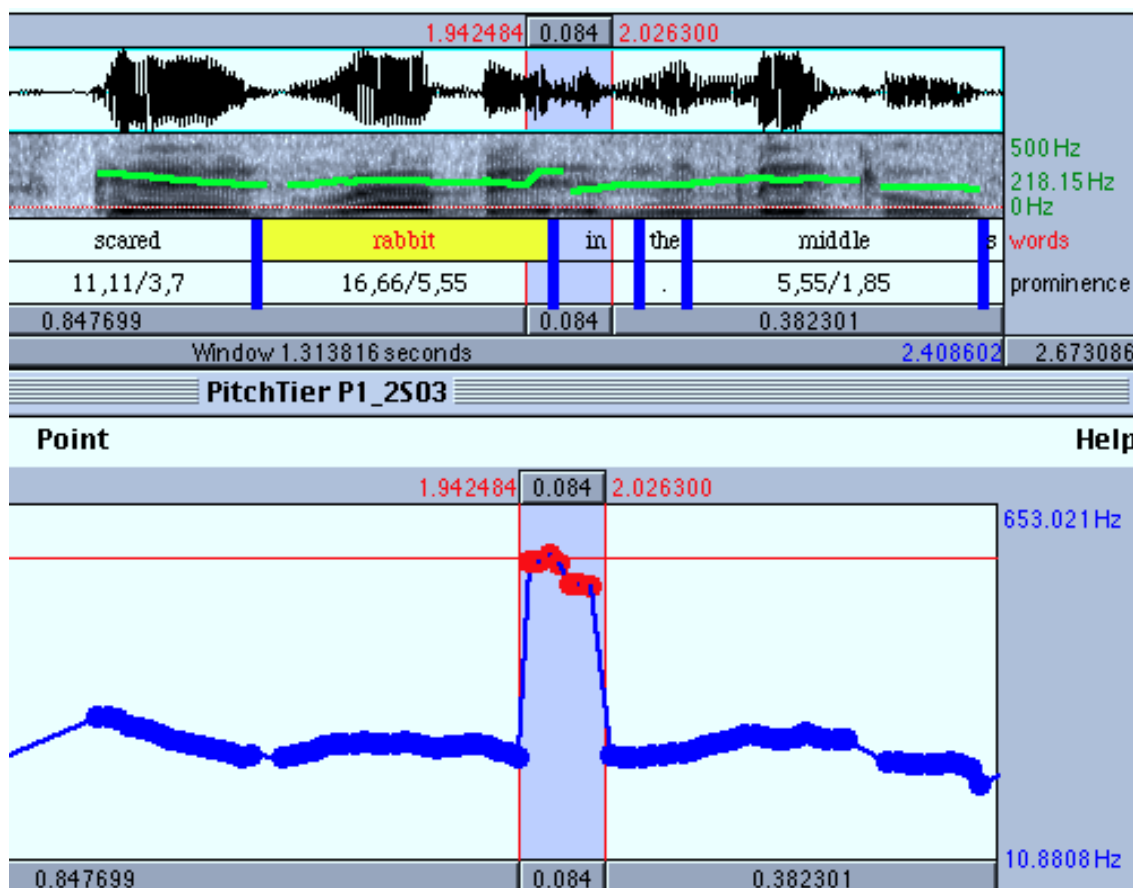


Figure 14 : P1.2S03, F0 brute

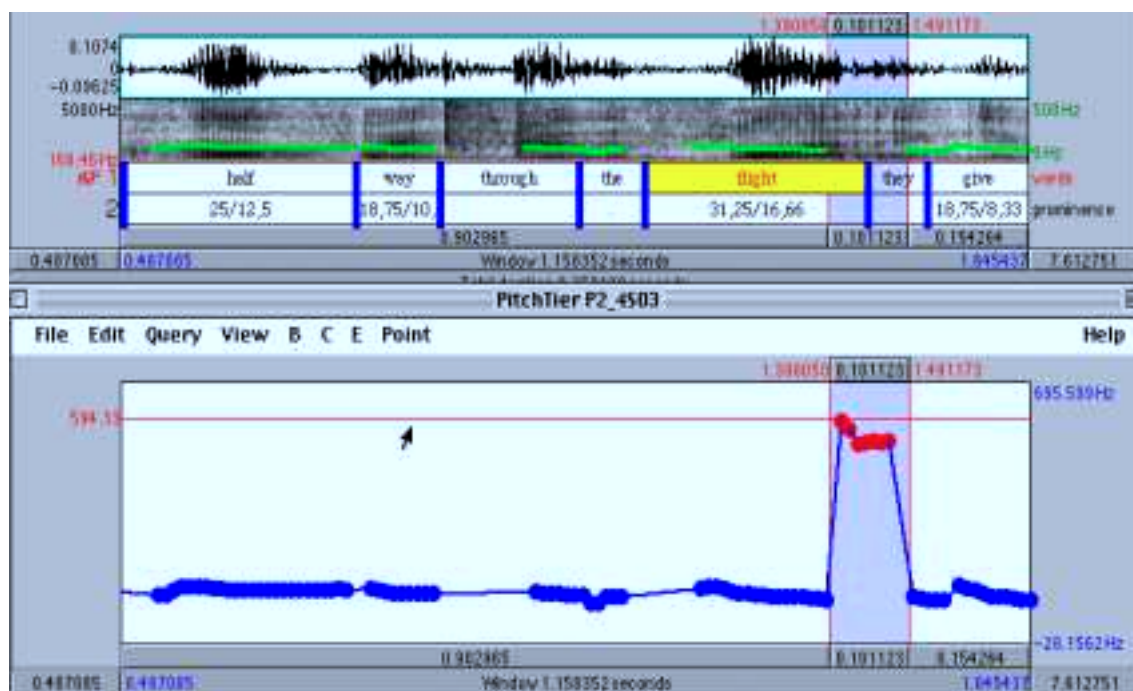


Figure 15 : P2.4S03, F0 brute

- la qualité de la voix fausse aussi la courbe : les valeurs sont très hautes quand la voix sonne un peu grinçante, comme dans P1.9S05 (☞ **exs.23**) : PM : *weapons* ; ou trop basses comme dans P2.3S01 (☞ **exs.24**) : C : *home*.

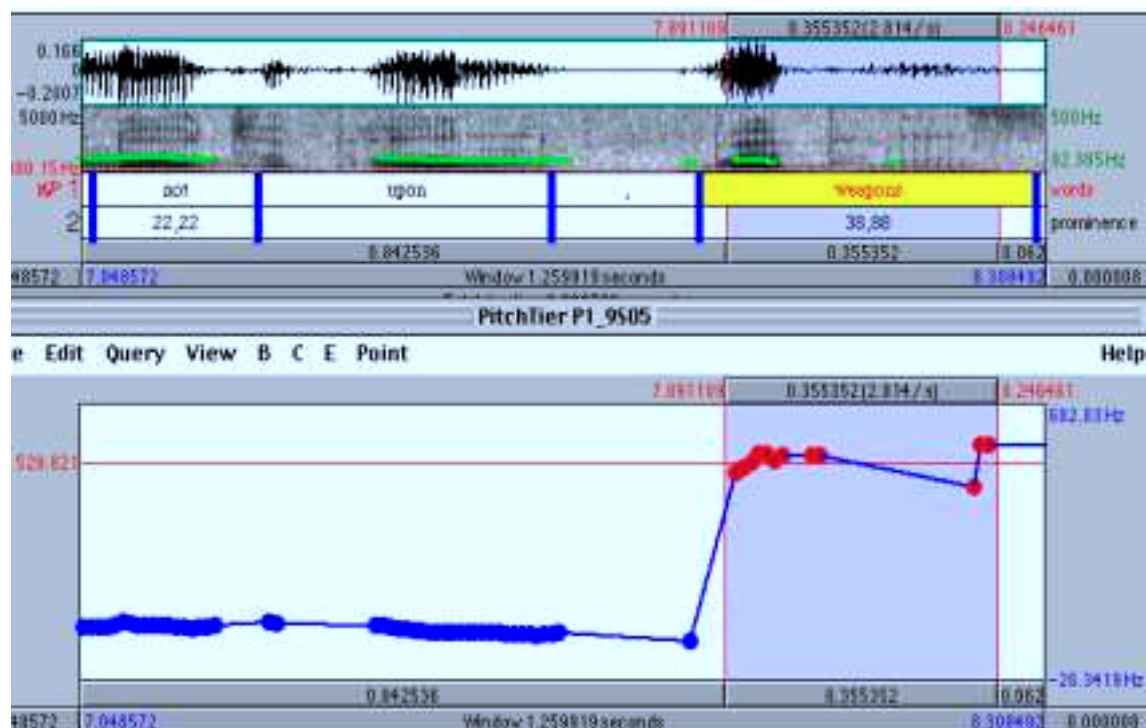


Figure 16 : P1.9S05, F0 brute

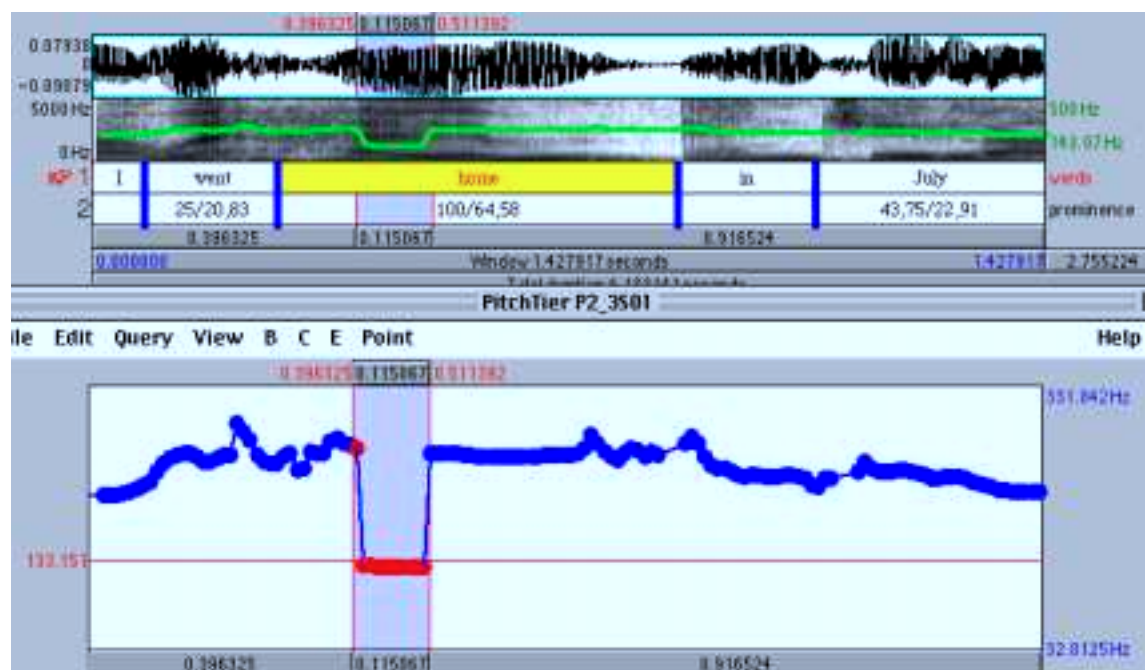


Figure 17 : P2.3S01, F0 brute

C'est à cause de ces variations micro-mélodiques que l'on n'étudie plus aujourd'hui la F0 brute, mais que des systèmes de stylisation de la courbe, gommant ces variations micro-mélodiques ont été développés : 't Hart (1984 : 195) explique qu'il faut essayer de tracer une ligne lissant la courbe de F0, qui serait plus représentative de ce que l'oreille perçoit :

[we should] try to draw a graph on top of the F0 curve which should represent the course of pitch as a function of time as it is perceived by the listener.

3.1.1.2 MOMEL (MOdelisation MELodique) et PRAAT

Un système de stylisation des courbes de F0 appelé MOMEL (MOdelisation MELodique) a été développé ces dernières années au Laboratoire Parole et Langage de l'Université d'Aix-en-Provence.

MOMEL (Hirst & Espesser, 1993) est un algorithme permettant de coder automatiquement des courbes de fréquence fondamentale au moyen d'une technique de régression quadratique modale asymétrique. MOMEL donne une représentation de la fréquence fondamentale par une séquence de points-cibles constituée par des couples de valeurs : <F0, temps>. Les points-cibles correspondent aux variations locales pertinentes de la courbe mélodique et permettent, interpolés par une fonction de type spline quadratique, de retrouver le profil supra segmental caractérisant globalement l'intonation.

L'algorithme MOMEL est multilingue (cf. Astésano *et al.*, 1997, pour une évaluation multilingue de cet algorithme et Herment-Dujardin, 1997, pour l'anglais), et l'idéal aurait été pour nous d'avoir recours à cet outil pour effectuer nos mesures.

Malheureusement, MOMEL n'est pas encore totalement compatible avec PRAAT. Il existe une version de PRAAT sur laquelle MOMEL a été implémenté et nous l'avons testée mais avons trouvé beaucoup trop d'erreurs. Les points-cibles sont trop espacés, ce qui fait que la courbe lissée est trop éloignée de la courbe d'origine et les résultats auraient été moins bons qu'avec la courbe de F0 brute. Cette version de PRAAT avec MOMEL ne fonctionnant pas sur la plateforme que nous utilisons, nous ne pouvons en montrer un exemple.

Le logiciel PRAAT (Boersma & Weenik, 1996), à partir duquel nous avons étiqueté notre corpus propose une version « lissée » (« smooth ») de la courbe de F0. Ce lissage est moins précis que celui que nous aurions obtenu grâce à MOMEL, car en effet il ne permet pas de gommer les variations micro-mélodiques, mais on peut effectuer des corrections manuelles, ce qui présente de gros avantages⁶².

Les figures 18 et 19 montrent des exemples de courbes de F0 brutes et lissées :

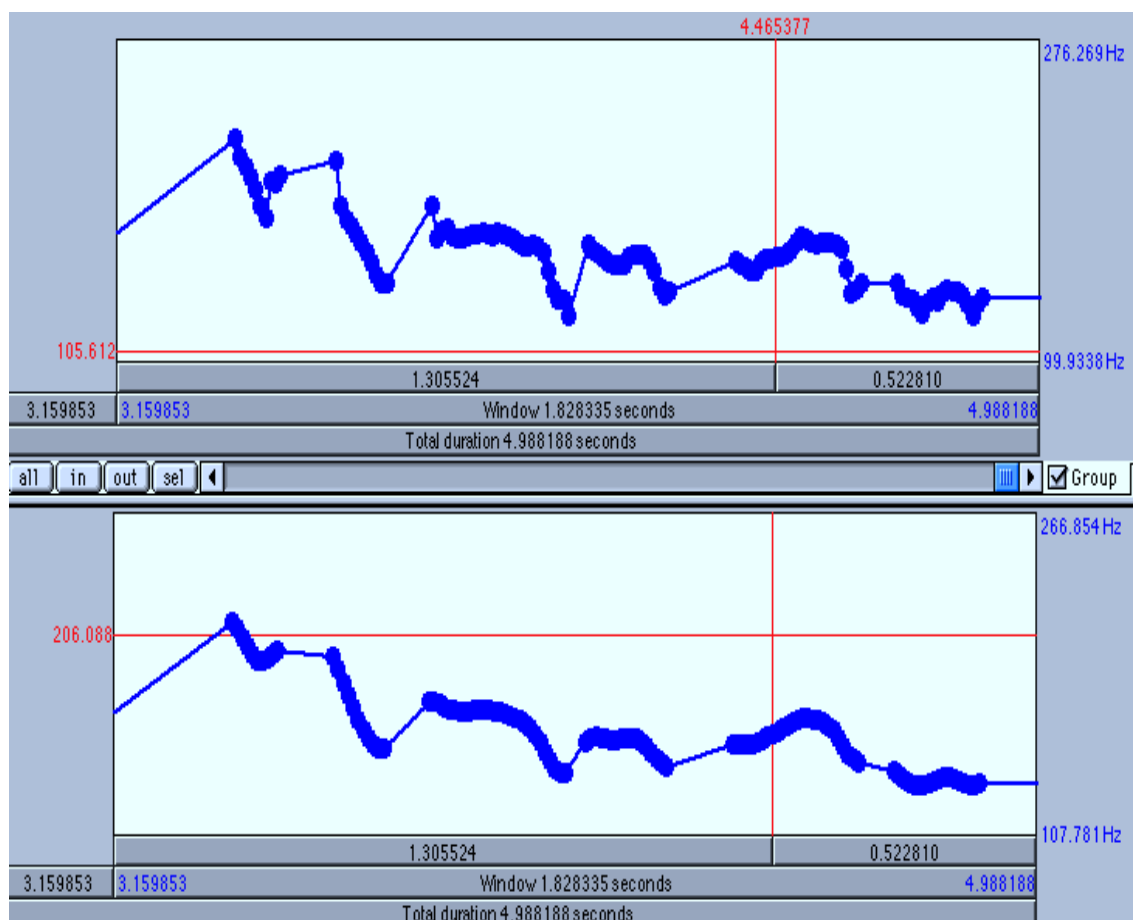


Figure 18 : P3.3S05, courbe brute en haut, courbe lissée en dessous

⁶² Cf. 3.1.1.3.

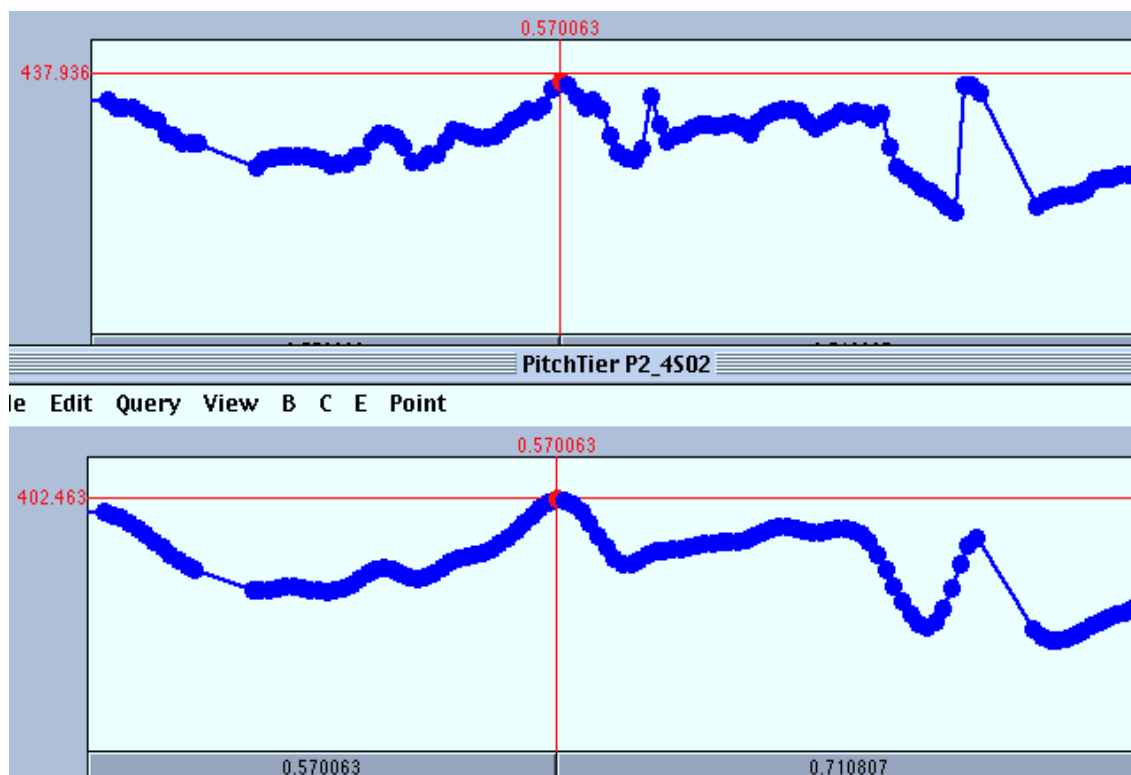


Figure 19 : P2.4S02, courbe brute en haut, courbe lissée en dessous

Nous avons donc effectué toutes nos mesures à partir de cette fonction « smooth » de PRAAT⁶³. Mais la version lissée ne corrige pas les erreurs pointées en 3.1.1.1, il faut donc pallier ce manque.

3.1.1.3 Corrections manuelles

PRAAT permet de sélectionner une fourchette de fréquences dans laquelle la courbe de F0 brute va s'inscrire. Ceci permet d'éviter les erreurs montrées dans les figures 13, 14, 15 et 16.

Prenons le locuteur PM par exemple. C'est un homme qui a une voix grave, il y a donc peu de chances qu'il monte jusqu'à 600 Hz dans l'aigu (600 Hz est la valeur par défaut donnée par PRAAT lorsque l'on ouvre la courbe de F0). Or on voit sur les quatre exemples ci-dessous (figure 20 : **exs.25, 26, 27, 28**) que l'on a bien des valeurs atteignant ce seuil. Sur les fenêtres montrées dans la figure 20, la ligne noire en haut est

⁶³ Tout au long de ce travail, nous avons bénéficié des nouvelles versions de PRAAT, que Paul Boersma ne cesse d'améliorer, et la dernière version que nous avons utilisée est PRAAT 3.9.19. Au jour où nous achevons ce travail, nous en sommes à la version 3.9.31.

l'intensité de voisement représentée par des chiffres. Les autres chiffres correspondent à chaque candidat possible de F0 auquel un poids est donné. Puis un algorithme de suivi fait un choix qui correspond aux points roses. On visualise donc la courbe de F0 grâce aux points roses et certains de ces points, que l'on trouve autour de 500 Hz dans les quatre exemples, semblent erronés. La solution pour éviter ces erreurs est de changer le plafond. On voit sur la figure 20, et c'est la raison pour laquelle nous montrons plusieurs exemples, qu'il semble typique chez PM d'avoir ainsi des valeurs beaucoup trop hautes, mais que sa hauteur mélodique maximale ne va pas au-delà de 350 Hz. Ces valeurs erronées sont peut-être dues à la qualité de la voix chez ce locuteur.

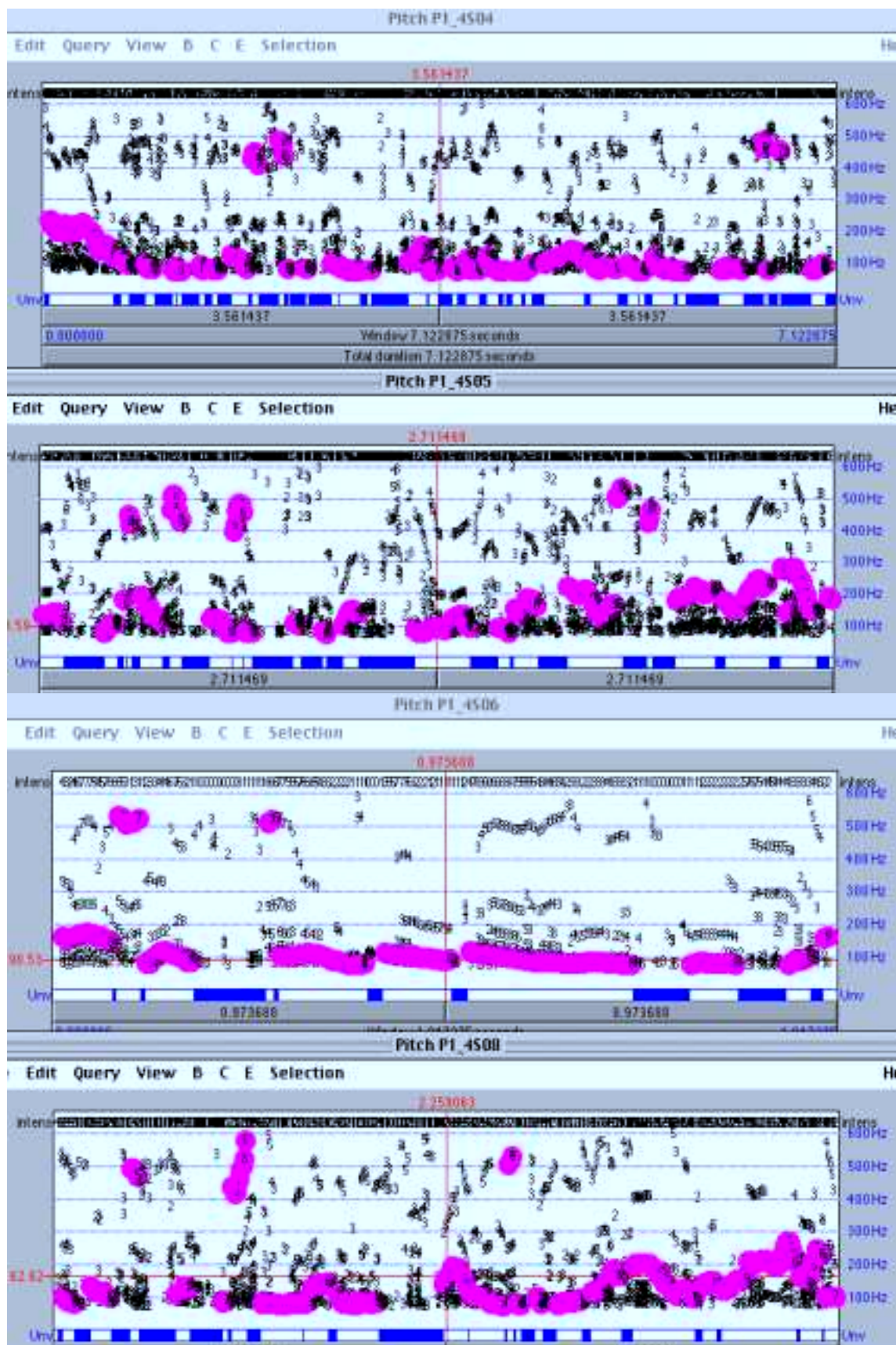


Figure 20 : courbes de F0 brute avec plafond à 600 Hz des segments P1.4S04, P1.4S05, P1.4S06 et P1.4S08, tous prononcés par PM.

Si l'on met le plafond à 350 Hz par exemple, au lieu de 600 Hz, toutes les valeurs erronées vont disparaître, comme le montre la figure 21 :

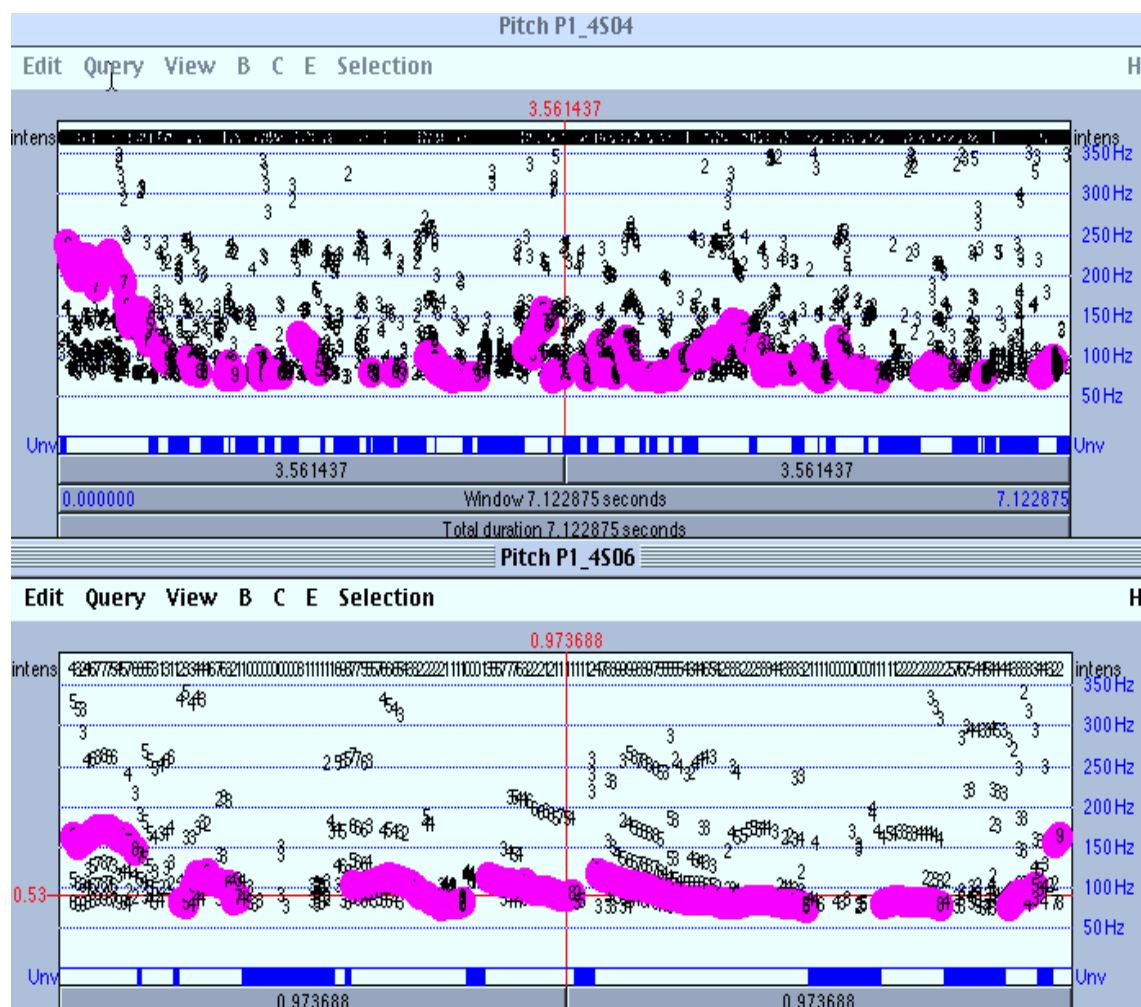


Figure 21 : courbes de F0 brutes avec plafond à 350 Hz pour les segments P1.4S04 et P1.4S06 prononcés par PM.

Cette manipulation peut être réalisée pour tous les segments et c'est donc un moyen d'éviter les erreurs dues aux variations micro mélodiques.

Qu'en est-il de l'erreur exemplifiée par la figure 17 ? De la même façon, il est possible de changer la valeur minimale de F0, c'est-à-dire le plancher. PRAAT donne 75 Hz par défaut, et nous avons repris le segment de la figure 17 et avons changé la valeur à 140 Hz. Voici les deux courbes avec le mot « home » sélectionné en bleu :



Figure 22 : courbes de F0 brute pour P2.3S01 :
la courbe du haut a un plancher à 75 Hz, la courbe du bas à 140Hz.

Ainsi pour chaque locuteur nous veillerons à ce genre de choses et tenterons d'éviter les valeurs erronées.

Voici un dernier exemple, le segment P1.3S01 (**exs.29**), prononcé par JT, un homme à la voix très grave, exemple qui montre toutes les étapes, à savoir la F0 brute avec les valeurs par défaut de PRAAT (75 Hz et 600 Hz), la F0 avec les valeurs adaptées (nous avons changé le plafond à 300 Hz), la courbe lissée (fonction « smooth ») et enfin la même courbe lissée mais visualisée différemment (fonction « pitchtier » : les points sont reliés par des lignes et la fourchette de valeurs s'adapte automatiquement, ce qui donne plus de relief) :

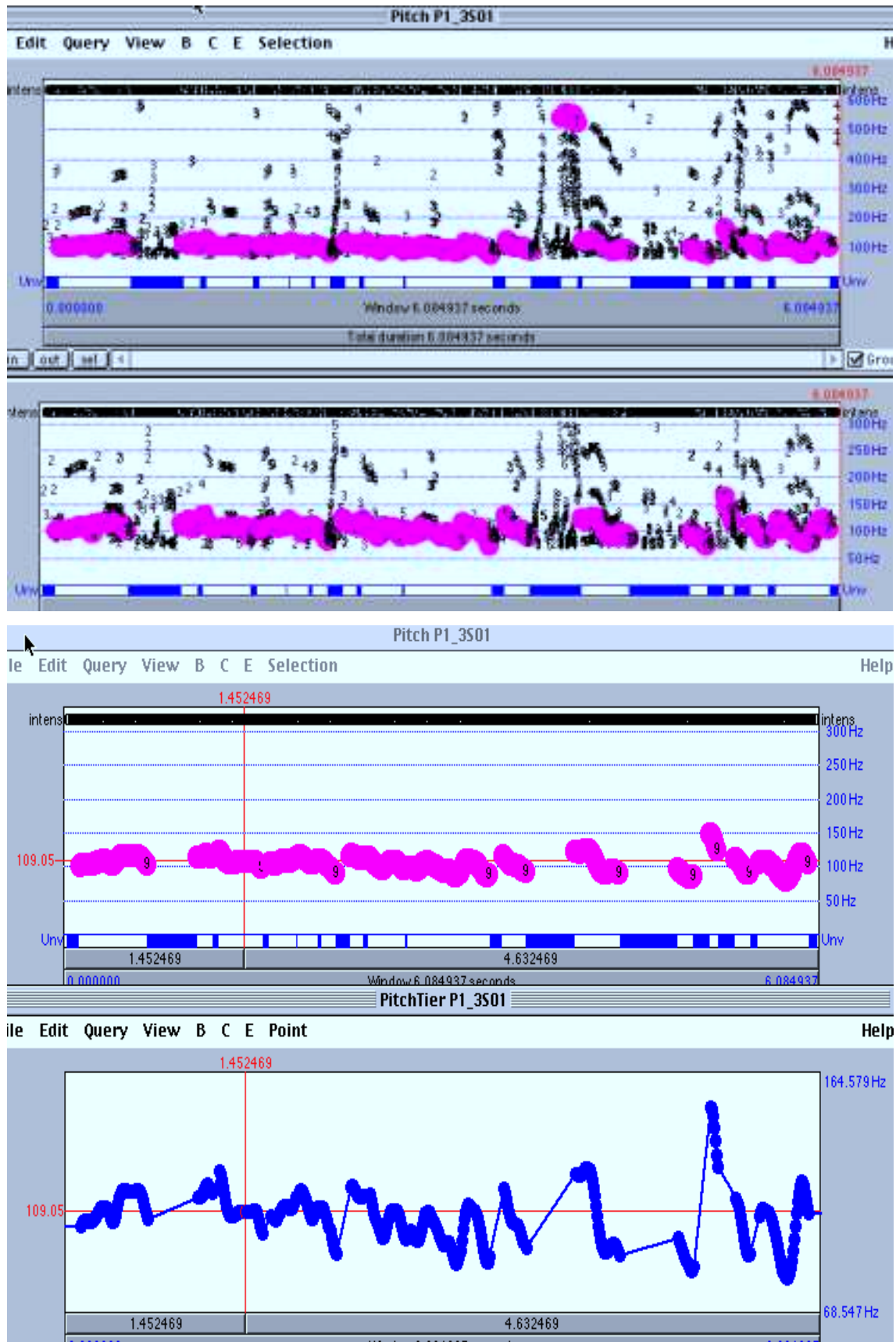


Figure 23 : segment P1.3501

3.1.2 Pré-analyses et données pertinentes

Avant d'effectuer les mesures précises à rentrer dans notre tableau de données, il était souhaitable de faire ce que nous appelons des pré-analyses afin de décider quels paramètres choisir. Nous avons effectué ces premières observations sur la courbe de F0 brute, cela nous permettant de repérer les segments où il y a des problèmes et de pouvoir compenser ensuite en corrigeant éventuellement la courbe.

Nous avons décidé de ne regarder pour ces pré-analyses que les segments dans lesquels au moins un mot était marqué emphatique par la plupart de nos auditeurs, c'est-à-dire par 88,88% d'entre eux dans le premier sondage (ce qui correspond à 16 auditeurs sur 18) et par 87,5% dans les deux autres sondages (14 auditeurs sur 16), ainsi que quelques passages les moins emphatiques possibles (pointés par peu ou pas d'auditeurs) pour chaque locuteur afin de pouvoir comparer les deux types d'énoncés et de courbes. Cette dernière étape a été très difficile car, comme nous l'avons déjà dit, nous avons sélectionné pour nos prétests des passages qui nous semblaient emphatiques et d'autres qui nous semblaient prononcés de façon tout à fait neutre, et nous nous sommes retrouvée confrontée au problème mentionné par Streefkerk *et al.* (1997), à savoir que même lorsque l'on donne des passages que l'on croit non emphatiques à écouter à des auditeurs, ils trouvent toujours quelques mots ou syllabes emphatiques : Streefkerk *et al.* ont mis en place une expérience testant l'importance de la hauteur mélodique. Les auditeurs entendent des phrases resynthétisées avec un contour intonatif monotone et doivent marquer les mots proéminents. Bien que la fréquence fondamentale soit reconnue comme le paramètre le plus important dans la perception de la proéminence, les auditeurs réussissent quand même à marquer des mots emphatiques. Streefkerk et son équipe en concluent que la fréquence fondamentale n'est pas le seul paramètre entrant en jeu dans la perception de l'emphase mais que d'autres corrélats acoustiques interviennent, comme la durée, et l'intensité (1997 : 110), mais nous nous demandons si cela n'est pas aussi dû au problème de la présentation et des instructions. Nous-même avons précisé dans les instructions des tests que tous les passages n'étaient pas forcément emphatiques mais au moins un mot dans tous les morceaux sélectionnés sans exception a été donné emphatique par au moins une personne.

3.1.2.1 F0 maximale et minimale

Ces pré-analyses nous ont permis de voir que les mots emphatiques à un haut degré⁶⁴ sont souvent porteurs d'un mouvement mélodique, et le plus souvent d'une chute :

- JT : 4 mots emphatiques, 4 chutes ;
- PM : 13 mots emphatiques, 8 chutes ;
- DD : 5 mots emphatiques, 4 chutes ;
- MM : 10 mots emphatiques, 6 chutes ;
- YB : 25 mots emphatiques, 19 chutes ;
- PS : 6 mots emphatiques, 3 chutes ;
- VB : 13 mots emphatiques, 5 chutes ;
- JC : 5 mots emphatiques, 5 chutes ;
- DL : 8 mots emphatiques, 3 chutes ;
- CB : 5 mots emphatiques, 1 chute ;

Il est intéressant de noter que dans la conversation, les locuteurs utilisent moins les chutes que dans les deux autres corpus.

Mais un mouvement mélodique n'est bien entendu pas toujours emphatique. Dans des phrases dites « neutres », le noyau du groupe intonatif reçoit en général un mouvement mélodique également, soit une chute, soit une montée, de la même façon que ce que nous avons observé pour nos mots emphatiques. L'intérêt ne réside donc pas dans le mouvement lui-même, mais bien dans l'amplitude de ce mouvement par rapport aux mêmes mouvements considérés non emphatiques. C'est la raison pour laquelle il semble nécessaire de mesurer le point où la fréquence fondamentale est maximale sur le mot choisi et aussi le point le plus bas sur le même mot. Ou faut-il prendre le mot qui suit ? Ou bien la même syllabe ?

3.1.2.2 Syllabe, mot ou pied ? Une alternative : l'unité tonale

Un problème théorique se pose dès à présent : faut-il analyser une seule syllabe, tout le mot, ou le pied⁶⁵ peut-être, comme certains linguistes le préconisent.

⁶⁴ Nous rappelons que nous avons calculé pour chaque mot un pourcentage d'emphase qui équivaut au pourcentage d'auditeurs ayant pointé ce mot comme emphatique. Nous considérons que ce pourcentage donne un degré d'emphase. Un mot fortement emphatique est un mot pointé par beaucoup d'auditeurs.

Nous devons d'abord préciser que nous considérons, à l'instar du plus grand nombre de linguistes, une unité supérieure que nous appelons unité intonative ou groupe intonatif (nous étudions ce point en détail ci-dessous, cf. 3.1.4).

Pour en revenir à notre problème, il semble impossible de n'analyser que la syllabe pour la simple raison qu'une chute est souvent perçue et produite d'une syllabe à l'autre dans un même mot (de la syllabe accentuée à la syllabe inaccentuée). Il est vrai que si le mot est monosyllabique, le mouvement mélodique a lieu sur la seule syllabe mais qui est aussi le mot. Nous ne prendrons donc pas la syllabe comme unité de base pour nos mesures de fréquence fondamentale.

Le mot semble être la solution la plus facile, mais nombre de linguistes utilisent comme unité rythmique le pied, à la suite d'Abercrombie (1964) qui définit le pied comme une syllabe accentuée suivie d'une séquence de syllabes non accentuées.

Jassem (1952) avait auparavant proposé un modèle sophistiqué de structure rythmique pour l'anglais en distinguant deux unités :

- l'unité rythmique étroite (« Narrow Rhythm Unit ») qui consiste en une séquence de syllabes contenant une syllabe accentuée suivie de syllabes non accentuées et
- l'anacrusse (« anacrusis ») qui est une suite de syllabes inaccentuées proclitiques.

La combinaison de l'anacrusse et de l'unité rythmique étroite constitue l'unité rythmique totale (« Total Rhythm Unit »).

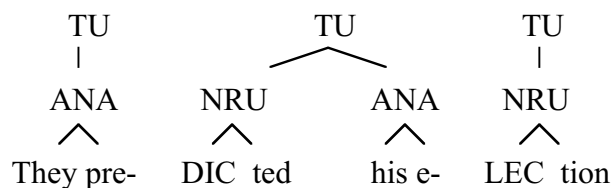
Au niveau mélodique, Jassem ne fait plus usage de l'unité rythmique totale mais regroupe au contraire l'anacrusse avec l'unité rythmique étroite qui la précède pour former une unité tonale (TU : « Tonal Unit »). Nous reprenons l'exemple de Hirst (1998 : 59) :

Rhythmic structure :



⁶⁵ Le pied est composé d'une syllabe accentuée suivie de toutes les syllabes inaccentuées jusqu'à la syllabe accentuée suivante. Un pied peut ne pas contenir de syllabe accentuée (en tête atone de groupe intonatif uniquement) ou pas de syllabes inaccentuées.

Melodic structure :



Contrairement à l'unité rythmique étroite, l'unité tonale est indépendante de la frontière de mot.

Abercrombie (1964) a repris les concepts de Jassem pour ne garder que l'unité tonale qu'il nomme le pied.

Halliday (1967a) a repris aussi ces idées et le pied, pour lui, est constitué d'une syllabe accentuée (« ictus ») suivie de syllabes inaccentuées (« remiss »), exactement comme pour Abercrombie.

Nous avons essayé de voir s'il serait intéressant d'appliquer le concept du pied pour nos analyses.

On remarque que dans le cas de mots longs, on trouve des tons creusés sur tout le mot :

☞ **exs.30** : P1.8S04 : *principe* : 408 – 137 – 257 Hz (mesurés sur la courbe de F0 brute, cf. figure 24 ci-dessous). Mais ici, le mot correspond aussi au pied, puisque la première syllabe du mot est accentuée.

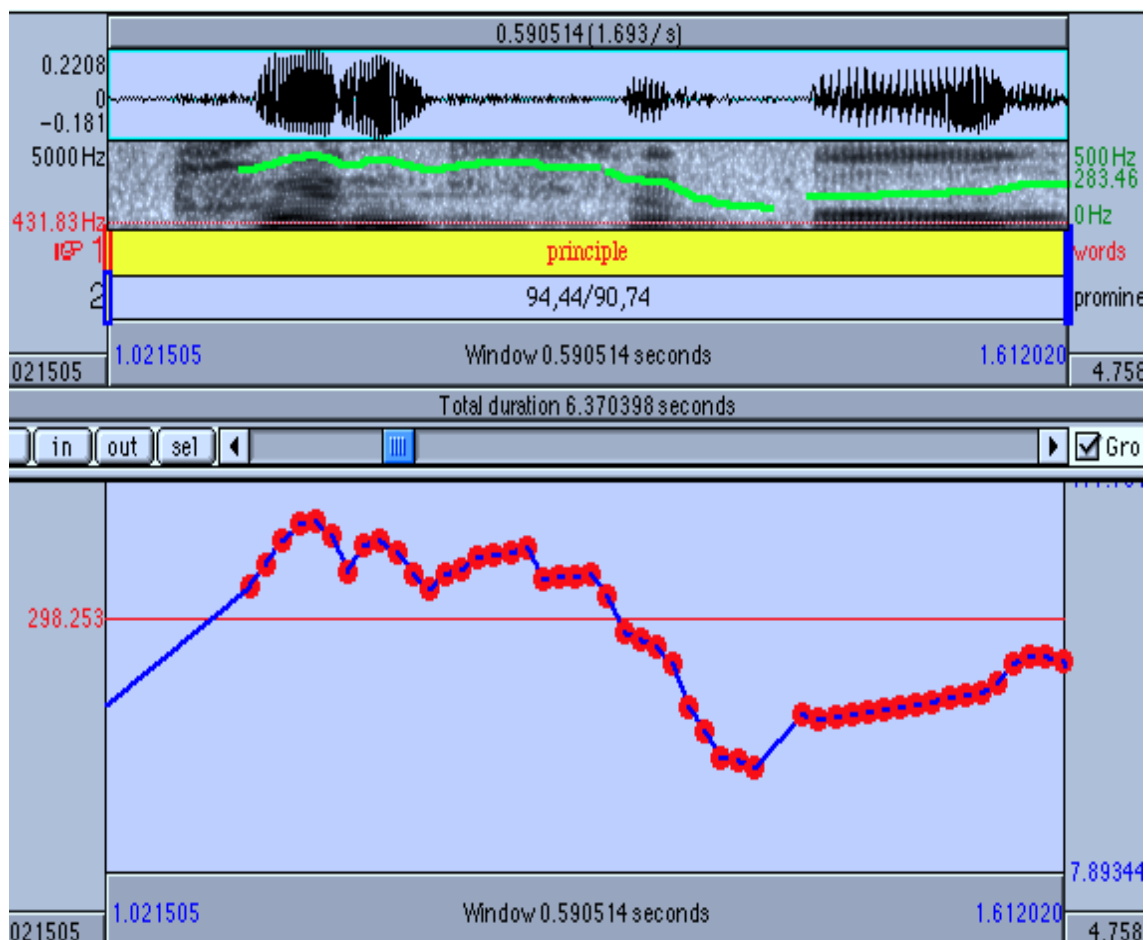


Figure 24 : P1.8S04 : ton creusé sur « principe »

Prenons donc un autre exemple : exs.31 : P3.1S09 : *extremely*. La deuxième syllabe est accentuée, et sur le mot entier on obtient les mesures suivantes : 180 Hz sur la 1^{ère} syllabe, puis rupture, on monte à 289 Hz et chute jusque 193 Hz sur « -tre- », puis statique sur la fin du mot (cf. figure 25).

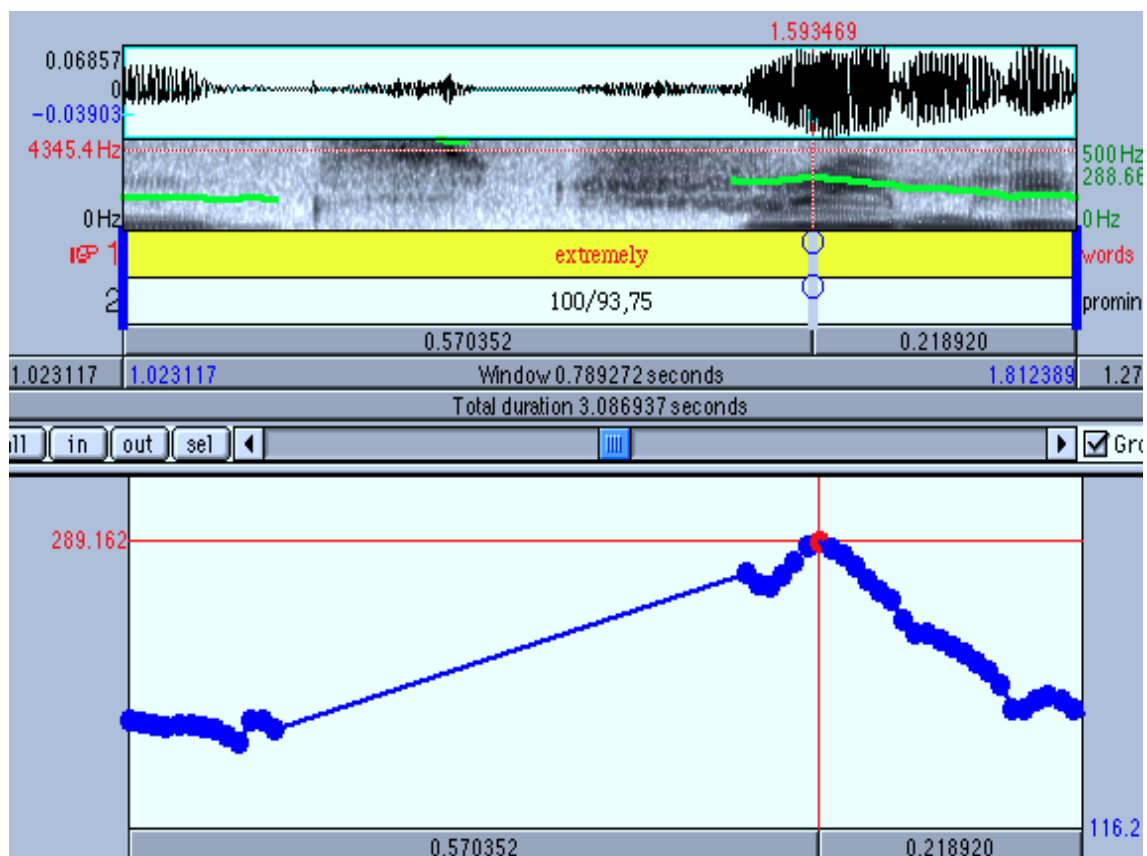


Figure 25 : P3.1S09

S'il n'y avait pas rupture entre la 1^{ère} et la 2^{ème} syllabe, on pourrait se dire que l'on a un schéma de type circonflexe (« rise-fall ») sur le mot entier alors que l'on a seulement une chute si l'on prend le pied. Mais il y a justement rupture et on parle de chute seulement, que l'on prenne le pied ou le mot. Nous avons regardé dans nos pré-analyses sur les mots très emphatiques uniquement les mots qui n'étaient pas accentués sur la première syllabe pour voir si une grosse différence se dégagait entre le mot pris dans son ensemble et le pied. Il se trouve qu'il n'y a que cinq mots dans ce cas : P1.3S07 (« remain »), P1.4S02 (« accounting »), P1.8S05 (« represent »), P3.1S01 (« arranged »), P3.3S03 (« authorities »). Pour « remain », par exemple (☞ **exs.33**), le début du mot est autour de 97 Hz, puis monte à 163 Hz sur la syllabe accentuée et redescend sur cette même syllabe à 134 Hz. La syllabe inaccentuée suivante (« in ») est à 88 Hz, puis un nouveau pied commence, en statique à 88 Hz également. La 1^{ère} syllabe « re- » est au même niveau que les syllabes précédentes. S'agit-il alors d'un ton circonflexe sur le mot entier ou simplement d'une chute sur le pied ? Notre oreille semble percevoir une chute uniquement. Il en va de même pour les autres mots. Nous

constatons que la ou les syllabe(s) précédant la syllabe accentuée du mot emphatique sont en fait dans la continuité des mots précédant le mot emphatique. C'est la raison pour laquelle nous penchons pour l'analyse du pied. Cependant, étant donné que nous avons décidé de mesurer le pic maximal de F0 et le pic minimal de F0, on aurait dans le cas de « remain », l'indication d'une chute de toute façon. Le problème n'est donc pas résolu entre le mot et le pied, même si nous sommes tentée de pencher pour le pied.

Une alternative intéressante s'offre à nous : certains linguistes font une combinaison de ces notions et distinguent une unité prosodique qui se situe au-dessus du pied et en-dessous de l'unité intonative : le mot prosodique ou phonologique (Selkirk, 1978, 1984 ; Nespor & Vogel, 1986), dont l'idée de base est le pied mais appliqué au niveau du mot plutôt qu'à celui de la syllabe.

Cette unité va être différente d'une langue à l'autre puisqu'elle est basée sur la structure rythmique de la langue. Hirst, Di Cristo & Espesser (2000) comparent le français et l'anglais, la première langue ayant une tendance à l'accentuation à droite, la seconde étant plutôt une langue à accentuation à gauche, et montrent les différences que cela entraîne : deux énoncés ayant la même structure syntaxique auront des structures prosodiques différentes :

a /'cup of/' tea

un 'verre / de 'vin

Dans les deux cas, on se base sur les accents lexicaux. Ainsi, cette unité supérieure au pied est constituée en anglais d'un mot accentué suivi d'une séquence de mots non accentués. L'unité intonative suivante, par exemple, sera découpée ainsi (les barres obliques marquant les frontières) :

'cause you / just / feel it in the / House of / Commons a / lot

Ce concept est très avantageux pour nous car il nous permet de ne pas perdre de vue l'importance de la structure rythmique et il nous facilite la tâche au niveau des manipulations. Mais cela pose un problème dans le cas d'un mot de plus d'une syllabe dont l'accent lexical ne tombe pas sur la première syllabe. On sera alors obligé de couper le mot. Or notre corpus est segmenté en mots et non pas en syllabes. Pourtant, en anglais la majorité des mots polysyllabiques sont accentués sur la première syllabe. Ils représentent environ 90% des mots. C'est pourquoi nous proposons une unité en quelque sorte dissidente mais inspirée de tous ces concepts. Nous garderons le nom

d'Unités Tonales à la suite de Hirst (1999) et Hirst *et al.* (2000) mais à la différence que nos unités tonales incluront les anacruses en début de mot, et seront totalement dépendantes des frontières de mot. La définition de nos unités tonales est donc la suivante : un mot accentué suivi de tous les mots inaccentués jusqu'au suivant. Si l'on reprend l'exemple ci-dessus

'cause you just feel it in the House of Commons a lot

le découpage sera absolument identique.

Ainsi nous avons des passages (P...) et des segments à l'intérieur de ces passages (proposés aux auditeurs pour les tests de perception : P...S...); à l'intérieur de ces segments on trouve des unités intonatives (P...S...UI...) et à un niveau encore inférieur, nous distinguons donc les Unités Tonales (P...S...UI...UT...).

3.1.2.3 Temps

Nous avons déterminé de quelle façon nous coupons nos énoncés, nous avons déjà vu qu'il fallait mesurer la valeur maximale de F0 ainsi que la valeur minimale de F0. Nous proposons de mesurer ces deux valeurs sur chacune de nos unités rythmiques, à savoir le mot, l'unité tonale et l'unité intonative. Nous pourrions ainsi comparer les occurrences emphatiques aux occurrences non emphatiques, les unités intonatives contenant un passage emphatique à celles qui n'en contiennent pas, etc...

Une fois ces deux valeurs trouvées ($MaxF0$ et $MinF0$), il est possible de calculer l'amplitude du mouvement mélodique, s'il y en a un bien sûr⁶⁶, sur le mot ou sur l'unité tonale. Pour l'unité intonative, il est difficile de parler de mouvement mélodique avec uniquement deux valeurs, mais le fait de les avoir mesurées nous permettra de voir si le mot emphatique par exemple est celui sur lequel la fréquence fondamentale est la plus élevée. Nous pourrions également voir si la F0 minimale correspond à un mot non emphatique ou pas.

Mais il faut aussi pouvoir dire si nous avons des montées ou des chutes. Le seul moyen de résoudre ce problème est de rajouter le paramètre temps en corrélation avec les maxima et minima de F0 ($TmaxF0$ et $TminF0$). Ainsi sur le mot ou sur l'unité tonale, si

⁶⁶ Nous y reviendrons dans la sous section 3.1.2.4.

le maximum se situe dans le temps avant le minimum, il s'agit forcément d'une chute et vice versa.

Pour avoir une idée plus précise de l'importance de ces mouvements mélodiques nous avons effectué quelques mesures et nous pouvons déjà dire, au vu de ces mesures, que pour la plupart des locuteurs les chutes sur les mots emphatiques sont plus grandes que les chutes sur les mots non emphatiques. D'autre part, les chutes emphatiques ont leur point de départ situé beaucoup plus haut dans l'aigu que les chutes non emphatiques.

Nous verrons si cette remarque est confirmée par les mesures faites sur tous les segments.

Nous avons beaucoup parlé des chutes, mais nous avons aussi observé quelques montées sur les mots emphatiques, ainsi que des tons creusés. Il n'y a aucun problème pour les montées grâce au facteur temps, par contre un problème se pose pour les tons creusés. Reprenons l'exemple du segment P1.8S04 (↵ **exs.30**), où le mot « principe » est emphatique et présente un ton creusé : 408 – 137 – 257 Hz (valeurs mesurées sur la courbe de F0 brute). Si l'on mesure sur ce mot (qui correspond aussi à l'unité tonale) la F0 maximale et la F0 minimale, on obtient 408 et 137, ce qui donne une chute, et la remontée n'est pas prise en compte. Comment pallier ce problème ? Nous avons regardé dans nos mots fortement emphatiques dans quels cas des tons creusés se produisaient. Il semble que ce soit toujours en fin d'unité tonale. Il est donc aisé de résoudre ce problème en rentrant le temps de la fin de l'unité tonale. Ainsi si la F0 minimale apparaît avant la fin, cela veut dire que l'on a une remontée. Nous devons donc ajouter le paramètre temps de la fin de l'UT. Pour encore plus de précision, il est même plus intéressant de rentrer le paramètre fin du mot à la place (*fin_mot*), la valeur pour le dernier mot de l'UT correspondant à la fin de l'UT. On pourra donc savoir si l'on a un contour sur l'unité tonale de type montant, descendant ou creusé (ou circonflexe de la même façon).

Mais nous avons vu qu'il ne fallait pas négliger le reste du groupe intonatif. Le contour général de tout le groupe est peut-être un paramètre déterminant dans l'emphase.

3.1.2.4 Contour intonatif général : pente, moyenne et écart-type

Il est important de savoir si ce qui entoure le mot emphatique est plutôt montant, ou descendant, ou ondulant (« sing song ») par exemple, plutôt dans le grave, l'aigu ou le médium. Comme le postulait déjà l'analyse classique de O'Connor & Arnold (1973), par exemple, et suivant les réflexions de Hirst (1987), Selting (1994), Uhmman (1992), nous pensons que cela peut être déterminant dans la perception de l'emphase. Une grande chute précédée d'un schéma de type « glide » dans la terminologie d'O'Connor & Arnold (1973), c'est-à-dire descendant progressivement jusqu'à la chute, ne sera pas forcément perçue emphatique, alors que la même chute précédée d'une montée préparatoire (type « high jump ») sera perçue emphatique. Cette hypothèse nous paraît tout à fait pertinente. De même nous avons vu que Hirst (1987) pense qu'il existe une différence dans la perception de l'emphase selon que l'on a une tonique de tête dans l'aigu (« high pre-head »), ou dans le grave (« low pre-head »)⁶⁷. Il nous faut donc trouver un moyen de rentrer dans notre tableau de données des paramètres qui permettent de rendre compte du contour tout entier. Nous avons déjà dit que nous allions mesurer la F0 maximale, la F0 minimale et le temps où apparaissent ces deux points pour l'unité intonative, ainsi que le temps de la fin de l'unité.

De ce fait, on pourra savoir si l'unité intonative tout entière est descendante ou plutôt plate par exemple. Mais il serait intéressant de calculer la moyenne de la fréquence fondamentale. Si le contour est plat avec de toutes petites variations, il est bon d'avoir la moyenne, d'autant que comme nous n'utilisons pas MOMEL, notre courbe risque de contenir des variations. Mais comment rendre compte d'un schéma ondulateur par exemple ? La moyenne ne sera pas révélatrice, et les F0 maximale et minimale nous donneront des indications faussées, on pourra croire que le schéma est de type montant ou descendant par exemple.

Il faut avoir recours pour cela au calcul de la pente, qui ne sera pas importante si on a une ligne mélodique ondulante (« sing song »). Mais comment distinguer grâce aux mesures un schéma ondulateur d'un schéma plat, pour lequel la pente ne sera pas élevée non plus ?

⁶⁷ Cf. 1.1.3.

Il faut alors ajouter un autre paramètre : l'écart-type⁶⁸. Il sera plus grand s'il s'agit d'une ligne mélodique ondulante que s'il s'agit de statiques. Ainsi nous proposons de rajouter ces paramètres, que nous mesurerons du fait pour les trois niveaux rythmiques que nous avons distingués.

3.1.3 Résumé partiel

Cette section nous a permis de déterminer quels paramètres mesurer et calculer en ce qui concerne la fréquence fondamentale.

Nos segments sont déjà découpés en mots et nous rajouterons deux niveaux rythmiques que sont l'unité tonale et l'unité intonative et nous mesurerons pour chaque niveau les valeurs suivantes :

- extremum où la F0 est minimale ($MinF0$) ; cette mesure est donnée en Hertz (Hz) ;
- temps de ce point ($TminF0$) ; cette mesure est donnée en secondes (s.) ;
- extremum où la F0 est maximale ($MaxF0$) ; cette mesure est donnée en Hertz (Hz) ;
- temps de ce pic ($TmaxF0$) ; cette mesure est donnée en secondes (s.) ;
- moyenne de la F0 de l'unité rythmique (mot, UT et UI) ($Mean$) ; cette mesure est donnée en demi-tons (dt) ;
- calcul de l'écart-type par rapport à la moyenne (SD ⁶⁹) ; cette mesure est donnée en demi-tons (dt) ;
- calcul de la pente de la courbe de F0 de l'unité rythmique (MAS ⁷⁰) ; cette mesure est donnée en demi-tons par seconde (dt/s) ;
- calcul de la pente sans saut d'octave ⁷¹ de la courbe de F0 de l'unité rythmique ($MASWOJ$ ⁷²) ; cette mesure est donnée en demi-tons par seconde (dt/s).

⁶⁸ La variance est la somme des écarts par rapport à la moyenne, élevés au carré. La variance donne donc une mesure en Hertz au carré pour la F0. L'écart-type est la racine carrée de la variance, il est donc exprimé en Hertz pour la F0.

⁶⁹ Standard Deviation.

⁷⁰ Mean Absolute Slope.

⁷¹ Le logiciel PRAAT propose de donner cette valeur, nous avons pensé qu'il serait peut-être intéressant de l'avoir étant donné que nous ne pouvons pas utiliser MOMEL pour lisser la courbe et gommer les variations micro-mélodiques.

⁷² Mean Absolute Slope Without Octave Jumps.

Nous souhaitons préciser que nous avons choisi de prendre comme unité de mesure les demi-tons pour les valeurs comme la moyenne et la pente ou l'écart type car il sera plus aisé de comparer les valeurs entre les hommes et les femmes d'une part et également parce qu'un écart-type en Hertz ne voudrait pas dire grand-chose, la perception des écarts en Hertz variant selon que l'on se trouve dans l'aigu ou dans le médium. Un écart en demi-tons correspond mieux à ce que l'on entend, l'échelle des demi-tons étant une échelle logarithmique.

Pour chaque mot de notre corpus primaire nous rentrerons donc les huit valeurs qui s'appliquent au mot, ainsi que les huit qui s'appliquent à l'unité tonale dans laquelle se trouve le mot, puis les huit qui s'appliquent à l'unité intonative dans laquelle se trouve le mot. Nous aurons ainsi pour commencer 24 valeurs en ce qui concerne la fréquence fondamentale pour chaque mot.

3.1.4 Découpage en unités tonales et intonatives

Nous avons donc dû au préalable découper notre corpus en unités tonales et en unités intonatives.

L'étiquetage en UT n'a pas vraiment posé de problème. Il est assez aisé de déterminer quels sont les mots accentués ou non, et quand nous avons un doute (dans le cas des auxiliaires ou des autres mots outils), nous nous sommes fiée à la prononciation du mot : si la prononciation était de forme réduite, le mot n'était pas accentué, et si le mot avait sa forme pleine, nous l'avons considéré accentué. On ne peut pas vraiment dégager de principe, et le contexte est fondamental :

Le contexte sous toutes ses formes est capital pour choisir les mots à accentuer. (Ginésy, 1995 : 196)

L'oreille est alors la seule à déterminer les mots accentués.

Mais le découpage en UI est parfois très délicat. Dans la littérature, les avis divergent d'ailleurs, à commencer par la terminologie.

3.1.4.1 Terminologie

Cette unité large que nous avons nommée Unité Intonative à la suite de Hirst (1998 : 59) prend différents noms selon les auteurs et surtout selon ce sur quoi elle est basée. Lieberman (1967), à l'instar de Sweet (1906) d'ailleurs, parle de « breath-groups » car il se base sur la dimension physiologique. D'autres parlent de « sense groups » (Kingdon, 1958b), se basant sur la dimension sémantique. Certains ont employé « tune » (Armstrong & Ward, 1926 ; Jassem, 1952), renvoyant plutôt à l'idée du contour intonatif, et distinguant différents « tunes ». De nombreux linguistes se basent plutôt sur des critères phonologiques, comme Crystal (1969) qui parle de « tone unit ». Sont encore employés « tone group » (Halliday, 1967a ; Gussenhoven, 1983), « intonation group » (Cruttenden, 1986), et même « intonation chunks » (Brown *et al.*, 1980).

3.1.4.2 Frontières d'unités intonatives

Lieberman (1967) se fonde sur la respiration pour définir ses groupes de souffle, arguant que la chute de la fréquence fondamentale sur la fin des énoncés s'explique par la baisse de pression de l'air subglottal. Dans les énoncés marqués, les muscles laryngés accroissent en tension et la fréquence fondamentale ne subit pas cette chute. Cette théorie s'est vue contredite par des études expérimentales (Ohala & Hirano, 1967) et ne rencontre plus d'adeptes aujourd'hui.

On a aussi souvent délimité les unités intonatives en fonction de critères grammaticaux ou sémantiques. Pour Kingdon (1958b : 162f), les groupes de sens sont définis comme des

[...] groups of words that have semantic and grammatical unity – not necessarily complete.

Selon Quirk *et al.* (1985 : 1598), les « tone units » sont des unités d'information.

On a aussi associé l'unité intonative à l'unité grammaticale, comme Halliday (1970 : 3) :

One clause is one tone group, unless there is good reason for it to be otherwise.

Enfin les linguistes phonologues comme Crystal (1969) se sont basés sur des critères phonologiques, dont quatre principaux :

les pauses, remplies ou non ;

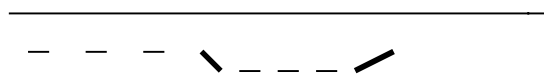
la présence d'une anacrusse qui indique généralement le début d'un groupe intonatif ;

l'allongement de la syllabe finale du groupe ;

un changement de direction de la mélodie ou un changement de hauteur mélodique (« step-up » ou « step-down ») non nucléaire, sur des syllabes inaccentuées souvent, qui signale en général la fin d'un groupe intonatif.

Malgré des réticences en ce qui concerne les pauses, qui peuvent marquer aussi l'hésitation ou la planification du reste du discours par exemple et pas forcément une frontière de groupe intonatif (cf. la sous-section 3.4 sur les pauses dans ce chapitre), ces indices sont désormais assez uniformément reconnus dans la littérature (cf. Cruttenden, 1986 : 36ff ; Couper-Kuhlen, 1986 : 75 ; Deschamps, Duchet, Fournier & O'Neil, 2000 : 185). Nous les avons appliqués aussi bien que possible pour procéder au découpage de notre corpus.

Il reste cependant difficile dans de nombreux cas de trancher, et des divergences persistent dans la littérature. Par exemple, là où Nicaise & Gray (1998) donnent deux groupes intonatifs, Ginésy (1995) n'en donne qu'un. Reprenons la phrase de Nicaise & Gray (1998 : 39) :



It would be awful if we failed

Lorsque cette phrase est prononcée avec une chute du médium dans le grave (qu'ils appellent « low fall ») sur « awful », et une montée du grave dans le médium (« low rise ») sur « failed », comme indiqué sur le schéma ci-dessus, Nicaise & Gray estiment que l'on a deux groupes intonatifs :

/ It would be awful / if we failed /

alors que Ginésy (1995 : 210-211) estime que l'on n'en a qu'un et il nomme ce schéma « dive ». On ne peut pas énoncer ou appliquer de règles bien précises dans ce cas, c'est le locuteur qui fait un choix et les deux options sont tout à fait possibles. Nicaise & Gray (1998 : 40) précisent d'ailleurs à propos de cette même phrase :

Ici, le choix de diviser en deux groupes intonatifs est un choix libre de la part du locuteur.

Nous pensons que si le locuteur opte pour deux groupes intonatifs, une pause est décelable entre « awful » et « if », et la proposition commençant par « if » est ce que l'on appelle parfois un « afterthought », d'où la petite pause, qui indique que l'on rajoute quelque chose. C'est ainsi que nous avons procédé et nous avons une fois de plus le sentiment qu'il est important de se baser sur du concret et sur de la parole spontanée pour nos analyses. Dans un cas comme celui cité ci-dessus, seul le son nous dira si l'on a un groupe ou deux.

Un autre problème peut se poser au niveau des mouvements mélodiques. Selon Nicaise & Gray, on reconnaît un groupe intonatif au fait qu'il contient une mélodie nucléaire. Selon Ginésy (communication personnelle), on peut avoir un « dive renforcé » par exemple, c'est-à-dire que la chute nucléaire est précédée par une première chute sur la tonique de tête ou dans le corps de l'énoncé. On aura quand même un seul groupe intonatif.

On voit ici toute la difficulté de découper le texte oral en unités intonatives, d'autant plus qu'il s'agit de discours spontané :

When we consider spontaneous speech (particularly conversation) any clear and obvious division into intonation-groups is not so apparent because of the broken nature of much spontaneous speech, including as it does hesitations, repetitions, false starts, incomplete sentences and sentences involving a grammatical caesura in the middle. (Cruttenden, 1986 : 36)

3.1.5 Résumé

Nous avons donc étiqueté notre corpus selon les principes décrits ci-dessus. Nous avons réécrit des fichiers TextGrid en ajoutant cette fois à l'étage des mots (nommé « words ») deux étages supplémentaires, l'un nommé « UT » pour Unités Tonales, et l'autre « UI » pour Unités Intonatives. Nous avons donc étiqueté nos 146 segments en UT et UI⁷³, que nous avons numérotées comme dans l'exemple suivant (☛ exs.33) :

⁷³ Ces fichiers TextGrid sont nommés TextGrid 2, cf. CD ROM : *Etiquetage : TextGrid 2*.

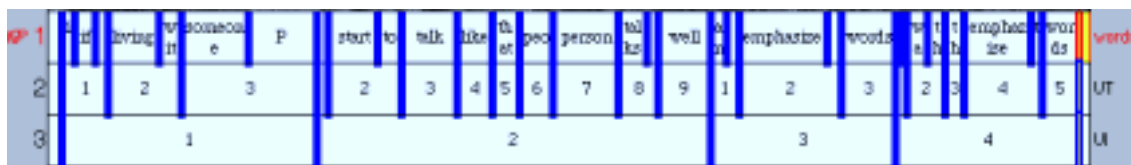


Figure 26 : P2.6S08 : TextGrid avec trois « tiers » : words, UT, UI.

Notre corpus était ainsi prêt pour que les mesures soient effectuées.

Au niveau pratique, c'est le logiciel PRAAT qui calcule toutes ces données. Nous les rentrons dans une feuille de calcul Excel afin de pouvoir ensuite imaginer des calculs pour trouver de nouveaux paramètres qui pourraient être pertinents⁷⁴ et y ajouter les paramètres en rapport avec l'intensité, les durées, etc... pour ensuite les transférer sur le logiciel de statistiques.

Afin d'alléger le travail, Daniel Hirst a écrit un script (*measure F0*)⁷⁵ qui demande à PRAAT de calculer pour chaque mot, chaque unité tonale et chaque unité intonative, les huit paramètres donnés ci-dessus (3.1.3), à savoir, la valeur minimale de F0 et le temps où elle apparaît, la valeur maximale de F0 et le temps à nouveau, la moyenne de F0, l'écart-type, la pente absolue et la pente absolue sans saut d'octave.

En résumé, nous aurons les valeurs maximale et minimale de F0, ce qui permet de savoir si l'on se situe plutôt dans le grave ou dans l'aigu. La pente nous permettra de savoir si le contour intonatif est plat ou non. Sur ce point, nous devons préciser que le logiciel PRAAT donne les valeurs absolues de la pente, ce qui ne nous permet pas de savoir s'il s'agit d'une montée ou d'une descente. L'écart-type nous renseigne sur l'ampleur du mouvement mélodique.

La pente et l'écart-type nous permettent donc de savoir à peu près quel type de contour nous avons :

- si la pente est faible et l'écart type élevé, nous avons un contour ondulatoire (type « sing song ») ou tout au moins plusieurs mouvements mélodiques qui s'annulent. Nous avons ce cas de figure en P2.3S08 (↩ **exs.34**), par exemple, où pour la partie « and you have to travel like half an hour at least to go to the beach », la pente sans saut d'octave est de 14,74 dt/s et l'écart-type est de 5,63 dt.

⁷⁴ Ces paramètres sont expliqués dans le chapitre 4.

⁷⁵ Cf. annexe 3a ou CD ROM : *Scripts : PRAAT : measure F0*.

- si la pente est importante et l'écart-type élevé, cela indique que l'on a un mouvement mélodique important, comme dans P1.4S03 (↩ **exs.35** dont la pente sur la première partie de la phrase est de 41,21dt/s et l'écart-type de 5,74 dt, ou encore dans la première partie de P1.6S02 (↩ **exs.36**) où la pente sans saut d'octave est de 60,04 dt/s et l'écart-type de 7,25 dt.
- si la pente est faible et l'écart-type faible aussi, il s'agit d'un schéma plat, comme dans P3.3S01 (↩ **exs.37**) où la pente pour le début du segment est de 14,71 dt/s et l'écart-type de 1,66 dt.

3.2 Durées

Nous avons choisi d'analyser le critère de durée, reconnu comme l'un des corrélats essentiels de la proéminence et de l'emphase. Nous avons vu dans le chapitre 1 que la plupart des études font état d'un allongement des mots emphatiques ou proéminents et aussi parfois d'un raccourcissement des syllabes ou mots adjacents.

Nous souhaitons donc mesurer la durée des mots donnés emphatiques dans notre corpus et également celle des mots adjacents aux mots emphatiques afin de voir si l'on peut appliquer ces critères à l'emphase. Un problème de méthode se pose alors : il faut choisir des durées moyennes pour chaque phonème, à partir desquelles on pourra dire s'il y a allongement ou raccourcissement. Comment faire pour déterminer ces durées moyennes ? Un rapide survol de la question s'impose afin de justifier de la méthode choisie.

3.2.1 Méthode

3.2.1.1 Niveau segmental ou supra segmental ?

Peterson et Lehiste (1960) ont été les premiers à introduire la notion de durée intrinsèque, définie comme la durée moyenne du noyau syllabique mesurée à partir de paires minimales, l'une avec une consonne voisée finale, l'autre avec une consonne non voisée finale. Ils ont trouvé en particulier que le noyau syllabique a tendance à être plus court s'il est suivi d'une consonne non voisée. Ils ont recueilli leurs données à partir de mots prononcés isolément.

De cette étude ont découlé plusieurs autres travaux dont Campbell (2000) fait la revue détaillée. Le problème central en ce qui concerne la durée des phonèmes est que celle-ci est sujette à de nombreux facteurs tels que la position dans l'unité intonative, la position dans le mot, l'accent lexical, la nature de la consonne qui suit quand il s'agit d'une voyelle, etc C'est la raison pour laquelle plusieurs linguistes ont tenté de donner des

modèles et des règles permettant de prédire les variations de la durée segmentale. Ils ont d'abord fait des mesures sur des mots lus isolément (House, 1961 ; Barnwell, 1971) puis se sont intéressés à de la parole plus spontanée (Klatt, 1975), trouvant d'ailleurs d'importantes différences entre les deux types de discours. Klatt (1979, 1987) a abouti à un modèle complet de prédiction des variations des durées segmentales basé sur onze règles en rapport avec la position, la nature des segments adjacents, etc ... Notons que parmi ces onze règles, Klatt donne une règle d'allongement dû à l'emphase : « lengthening for emphasis ».

Crystal & House (1988) ont comparé des lectures rapides et lentes et ont essayé de vérifier les règles données dans la littérature. Les résultats ne vont pas dans le sens de ces règles d'allongement ou de raccourcissement, à l'exception d'une ou deux. Mais ce que les travaux de Crystal & House mettent en évidence, c'est que les groupements segmentaux semblent avoir une durée moyenne à peu près équivalente (autour de 50 ms).

D'autres modèles de prédiction des durées segmentales mettent en avant la régularité rythmique caractéristique du discours fluide et donc s'intéressent à un niveau supra segmental et non plus segmental.

Certains se placent au niveau du pied (« foot »), comme Witten (1977) ou encore Young & Fallside (1980), d'autres au niveau de l'accent (« stress ») (Coker *et al.*, 1973), d'autres encore font des études statistiques (Riley, 1992 ; Pitrelli, 1990). Toutes ces études se basent néanmoins sur l'idée que chaque phonème a une durée intrinsèque stable, sujette uniquement à certains facteurs liés à l'environnement.

3.2.1.2 Niveau supra segmental : Campbell

Campbell (2000 : 302) pense pourtant qu'il n'y a pas vraiment de durée intrinsèque stable pour chaque phonème :

speech sounds have only minimal 'inherent' duration characteristics (if at all).

Il démontre, figures et chiffres à l'appui, que non seulement on observe des différences considérables dans les durées des phonèmes en contextes identiques entre différents

locuteurs, mais qu'en plus on observe des variations dues au style de parole (entre de la lecture et de la parole spontanée par exemple). D'où la difficulté de prédire les durées :

there is so much variation in the natural durations that we can not easily determine a 'right' duration for any context. (2000 : 304)

Et pourtant, quand on écoute de la parole synthétique, il est facile de pointer à quel endroit il y a une 'mauvaise' durée. Campbell & Isard (1991) ont donc proposé un modèle basé sur la syllabe, « the syllable-timing hypothesis » (STH). Un des principes fondamentaux de cette théorie est qu'à la fois les consonnes et les voyelles ont des caractéristiques d'allongement qui sont similaires. La STH tente d'expliquer comment tous les segments d'une syllabe peuvent partager l'espace disponible selon leur élasticité propre. Et en effet s'ils trouvent des variations considérables dans leur corpus au niveau des phonèmes, ces différences sont grandement atténuées lorsqu'ils se placent au niveau de la syllabe. Il découle de cela qu'il est plus fiable de se placer au niveau de la syllabe quand on parle de prédiction :

because there is less speaker-specific variation in the durations at the syllable level, we can consider them to be reliable as a basis for prediction. (2000 : 309)

La force de la théorie de Campbell est qu'il se place en fait sur deux niveaux grâce à deux concepts-clé : l'élasticité (« elasticity ») et l'ajustement (« accomodation »). Pour résumer les choses simplement, un ajustement a lieu au niveau de la syllabe grâce à l'élasticité de chaque segment. En d'autres termes, c'est au niveau de la durée de la syllabe que les durées des segments sont ajustées. La théorie de Campbell permet de prédire la durée des phonèmes (cf. Campbell, 1992a) à partir d'un niveau suprasegmental.

Grâce à cette théorie, Campbell (1992b) a pu extraire des durées moyennes pour chaque phonème de l'anglais.

3.2.2 Paramètres

Nous avons décidé d'utiliser les données de Campbell pour effectuer notre travail sur les durées. Ces données permettent de prédire les durées moyennes des mots de notre

corpus. Nous pourrions ensuite comparer ces durées prédites à nos durées réelles et voir s'il y a un allongement ou un raccourcissement de nos mots par rapport aux prédictions. N. Campbell a laissé ces précieuses données à la pointe de la recherche à notre disposition lors de sa visite au Laboratoire Parole et Langage d'Aix-en-Provence en décembre 2000.

A partir de notre corpus étiqueté en mots⁷⁶ et en phonétique⁷⁷, et grâce à un script (*mtostat*)⁷⁸, écrit par Cyril Auran (doctorat en cours), nous avons pu réaliser ces calculs et nous les intégrerons dans notre tableau de données.

Nous aurons donc les paramètres suivants pour la durée :

- la durée réelle du mot en millisecondes (*dur_reelle_mot*) ;
- la durée prédite du mot en millisecondes (*dur_predite_mot*) ;
- l'allongement (ou raccourcissement si l'on a une polarité négative) en millisecondes (*allong_ms_mot*) ;
- l'allongement (ou raccourcissement si l'on a une polarité négative) en pourcentage (*allong_%_mot*).

⁷⁶ Cf. 2.3.2.

⁷⁷ Cf. 2.3.3.

⁷⁸ Cf. annexe 3c ou CD ROM : *Scripts* : *PERL* : *mtostat*.

3.3 *Intensité*

3.3.1 Intensité globale, fiabilité et parole naturelle

L'intensité est également un paramètre reconnu pertinent par la plupart des linguistes dans la perception de la proéminence⁷⁹.

Cependant les études montrant de façon indéniable l'importance de l'intensité sont des études basées sur de la parole de laboratoire, pour lesquelles les enregistrements sont effectués dans des conditions optimales. Car en effet lorsque l'on s'intéresse à l'intensité, un problème crucial surgit : comment interpréter des données sur l'intensité globale alors que l'on sait que si le locuteur tourne la tête par exemple ou passe son bras devant sa bouche, l'intensité va varier. De même l'intensité donnera des mesures différentes en fonction de la qualité des micros utilisés. Des données dépendantes de tant de variables semblent donc inaptes à être interprétées :

[...] intensity variation will never have communicative significance for the simple reason that intensity is too susceptible to noise. If the speaker accidentally turns his head or passes a hand before his mouth, intensity drops of greater magnitude than those caused by the difference between stressed and unstressed syllables will easily occur. For this reason, manipulating intensity in stress perception experiments seems ill-advised. (Sluitjer & van Heuven, 1996 : 2372)

C'est la raison pour laquelle les travaux qui traitent de façon précise de l'intensité se basent sur des enregistrements faits de façon draconienne. Paradoxalement, les deux chercheurs cités ci-dessus analysent quand-même l'intensité globale dans leurs travaux mais prennent toutes les précautions possibles⁸⁰.

Alors comment faire quand il s'agit de parole spontanée ?

⁷⁹ Cf. 1.2.2.1.

⁸⁰ Cf. 1.2.2.3 : les enregistrements sont réalisés en laboratoire avec du matériel très perfectionné et la tête des locuteurs est bandée à un appuie-tête afin qu'ils ne puissent pas bouger et que la distance bouche-micro reste constante.

Sluitjer et van Heuven se sont intéressés à la balance spectrale en relation avec la perception de l'accent lexical, et nomment d'ailleurs un de leurs articles « intensity revisited » (1993).

Les études de Sluijter & van Heuven (1993, 1995) ainsi que les travaux de Campbell (1995) et Campbell & Beckman (1995)⁸¹ nous ont poussée à nous intéresser à ce point. Nous aurions aimé intégrer dans nos paramètres des données sur la balance spectrale, mais cela est encore trop récent et n'a pas été assez largement étudié. Nous ne disposons pas des outils nécessaires à une telle analyse et n'avons pas pu y avoir recours.

Nous nous trouvions alors dans une impasse, l'analyse de l'intensité globale étant d'autant moins fiable que nous travaillons sur de la parole spontanée.

Nous estimons pourtant qu'il est impossible de ne pas prendre l'intensité en compte, notre étude présenterait alors une lacune trop importante. Nous avons donc fait des pré-analyses pour ce paramètre aussi, souhaitant vérifier si les données ne risquaient pas d'être complètement faussées.

3.3.2 Pré-analyses

A nouveau grâce à PRAAT (Boersma & Weenik, 1996), nous avons la possibilité de visualiser une courbe d'intensité (« intensity tier »), qui, selon le même principe que la courbe « pitchtier » de F0, donne les points d'intensité reliés par des lignes, afin de mieux visualiser. Nous avons regardé ces courbes pour de nombreux segments et avons effectué des mesures dans le but de voir si l'on trouvait des constantes.

Nous avons procédé par locuteur, pour pouvoir effectuer des comparaisons d'un segment à l'autre et à notre étonnement, les valeurs d'intensité semblent tout à fait pertinentes et non faussées, si l'on ne perd pas de vue que l'on se trouve en discours spontané bien évidemment. Nous allons revenir sur ce point.

Nous avons mesuré la valeur maximale et la valeur minimale de plusieurs segments d'un même locuteur et avons mesuré des mots très emphatiques et des mots peu emphatiques. Nous avons regardé quelle était l'intensité des voyelles inaccentuées et des voyelles accentuées afin d'avoir une bonne vue d'ensemble. Nous avons également

⁸¹ Cf. 1.2.2.3 pour plus de détails sur ces travaux.

mesuré l'intensité des consonnes sourdes et des voyelles réduites, mais cela est bien moins pertinent car comme leur nom l'indique les consonnes sourdes sont prononcées sans voix, il est donc inadapté de mesurer l'intensité d'un souffle d'air, et les voyelles réduites sont souvent très courtes et difficiles à mesurer⁸².

Pour le locuteur DD par exemple, nous avons analysé sept passages, et les résultats sont les suivants :

- les valeurs les plus élevées se situent entre 77,75 dB et 73,8 dB. Mais dans trois segments, une personne parle en même temps que DD, il hausse donc le ton pour gagner la parole et se faire entendre. Si l'on enlève ces trois valeurs (77,75 ; 76,4 et 76,75 dB), la valeur la plus haute est 75,9 dB ;
- les valeurs les moins élevées pour les sept segments se situent entre 37 et 43 dB ;
- pour les consonnes sourdes, les valeurs vont de 40 à 42 dB ;
- pour les voyelles réduites, on se situe entre 58 et 62 dB ;
- les voyelles inaccentuées sont autour de 70 dB (entre 68 et 72 dB) ;
- enfin on trouve des valeurs allant de 73 à 75 dB pour les voyelles accentuées.

Le plus gros écart n'est que de 6 dB (et il s'agit des valeurs minimales, qui nous importent moins *a priori*) pour sept segments. Cela nous paraît peu et donc encourageant.

Nous avons donc continué pour d'autres locuteurs et avons mesuré 19 segments pour MM, ce qui est révélateur. On en arrive aux mêmes conclusions :

- valeurs les moins élevées : de 38 à 43,3 dB ;
- valeurs les plus élevées : de 78,2 à 73 dB ;
- consonnes sourdes : de 39 à 46 dB ;
- voyelles réduites : de 37 à 70 dB⁸³ ;
- voyelles inaccentuées : de 61 dB (un seul passage, sinon 67 dB) à 71 dB ;
- voyelles accentuées : de 72 à 74 dB.

Au vu de ces résultats nous avons encore analysé deux autres locuteurs (JT et VB)⁸⁴ et en avons tiré quelques conclusions.

⁸² Les détails des pré-analyses et des mesures prises sont donnés en annexe 3d.

⁸³ Ici l'écart est important mais cela présente peu d'intérêt, comme nous l'avons dit.

⁸⁴ Cf. les détails en annexe 3d.

On note une tendance au déclin de l'intensité tout au long de la phrase chez VB mais pas aussi nettement chez les autres locuteurs. Peut-être cela est-il dû au type de discours et nous rejoignons ce que nous avons déjà dit à propos de la fréquence fondamentale. Il faudra éclaircir ce point et rentrer dans notre tableau de données le paramètre type de discours.

Le mot le plus fort en intensité est souvent celui qui comprend la tonique de tête du groupe intonatif et pas toujours le mot emphatique, ou bien souvent, les deux, la tonique de tête et le mot très emphatique sont à peu près au même niveau d'intensité. Prenons l'exemple de JT (↵ exs.38 : P1.3S05 et ↵ exs.39 : P1.3S08) :

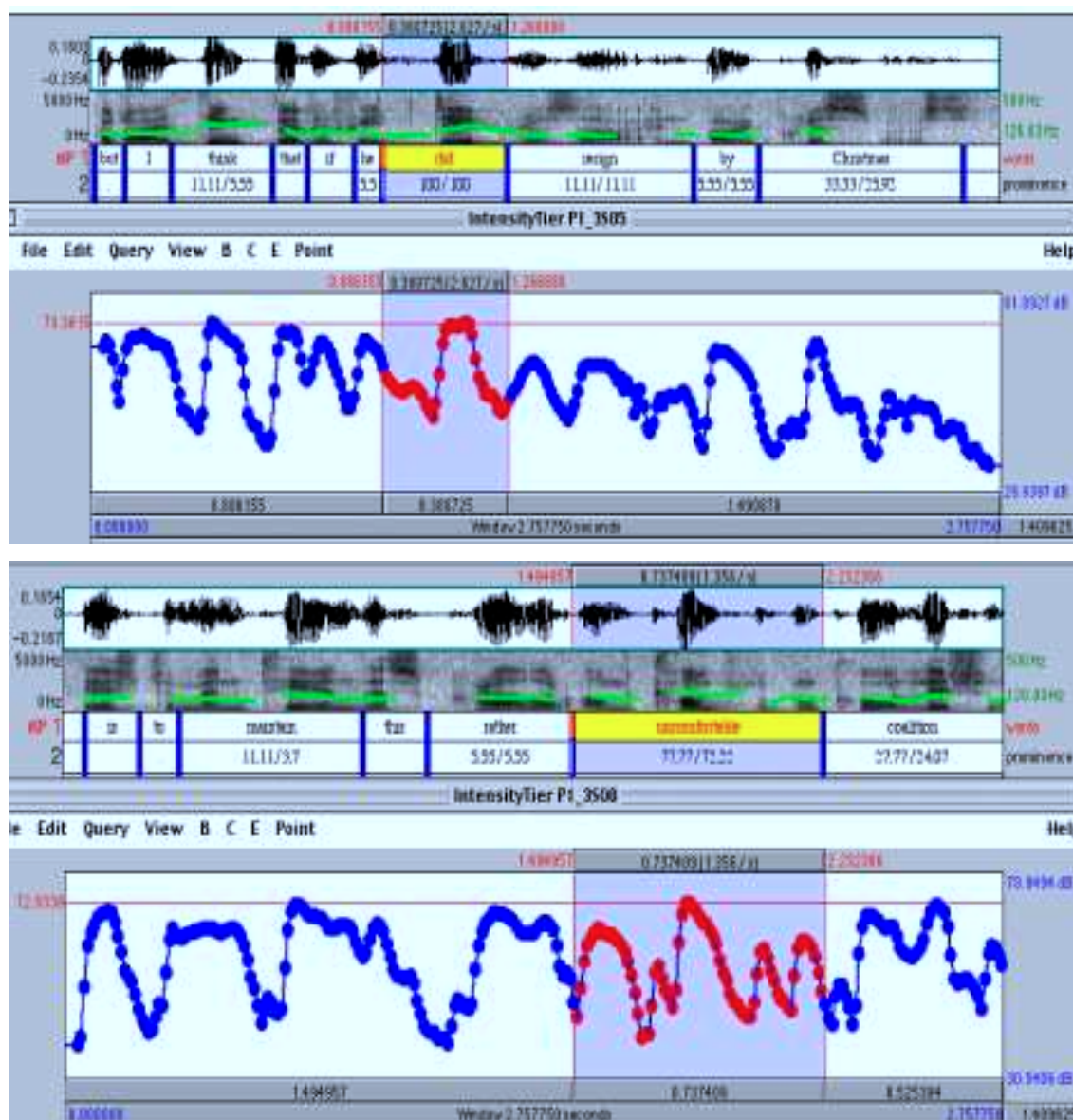


Figure 27 : P1.3S05 et P1.3S08 : son, TextGrid et « intensity tier »

Dans la fenêtre « intensity tier », le chiffre en bleu en haut à droite représente le plafond et le chiffre en bleu en bas à droite le plancher. Le chiffre en rouge à gauche donne la valeur la plus élevée du segment (nous avons sélectionné manuellement le point le plus haut). On voit sur cette figure que l'intensité correspond à l'emphase mais que les toniques de tête (« think » et la syllabe accentuée de « maintenant ») sont également fortes en intensité.

Face à ces premières conclusions, nous avons été confortée dans l'idée qu'il fallait mesurer l'intensité, mais de façon très systématique, et nous ne perdrons pas de vue lors de l'interprétation des résultats les réserves que nous avons émises au début de cette sous-section.

3.3.3 Intensité globale et paramètres

Pour tenter de pallier le problème décrit ci-dessus, la seule solution consistait à mesurer l'intensité sur des points très rapprochés, afin de pouvoir comparer des segments proches, cela minimisant le problème de la variabilité.

Nous avons repris les unités tonales déterminées pour l'analyse de la fréquence fondamentale⁸⁵ et avons décidé de mesurer l'intensité de chaque unité tonale, pour avoir le plus de précisions possibles. Nous ne mesurons pas l'intensité pour chaque mot car celle-ci ne peut être déterminée par PRAAT pour un segment inférieur à 64 millisecondes, trop de mots n'auraient donc pas été mesurés (pour la F0, PRAAT mesure à partir de 40 ms).

Nous avons repris les mêmes idées que pour l'analyse de la fréquence fondamentale : tout comme nous avons mesuré le pic de F0 maximale et la F0 minimale, il nous a semblé nécessaire de mesurer le pic d'intensité, c'est-à-dire l'intensité maximale et l'endroit où elle se trouve, ainsi que l'intensité minimale et le lieu correspondant. Nous avons donc quatre paramètres :

- l'intensité minimale de l'unité tonale (*MinInt_UT*) mesurée en décibels (dB) ;
- le temps de l'intensité minimale (*TMinInt_UT*) mesurée en secondes (s.) ;

⁸⁵ Cf. 3.1.2.2.

- l'intensité maximale de l'unité tonale (*MaxInt_UT*) mesurée en décibels (dB) ;
- le temps de l'intensité maximale (*TMaxInt_UT*) mesurée en secondes (s.) ;

auxquels nous avons rajouté :

- la moyenne de l'intensité de l'unité tonale (*MeanInt_UT*), mesurée en décibels (dB) et
- l'écart-type par rapport à la moyenne (*SDInt_UT*), donné en décibels (dB), qui peut être un indice selon les mots que l'on trouve dans l'unité tonale et le degré d'emphase.

Cependant nous avons estimé qu'il serait aussi peut-être pertinent dans les cas où l'intensité ne semble pas être sujette à des variations dues à des bruits ou mouvements extérieurs, de prendre les mêmes mesures sur tout le groupe intonatif - l'UI - (selon le même principe que pour la fréquence fondamentale) :

- *MinInt_UI* ;
- *TMinInt_UI* ;
- *MaxInt_UI* ;
- *TmaxInt_UI* ;
- *MeanInt_UI* ;
- *SDInt_UI*.

Comme pour la fréquence fondamentale, le logiciel PRAAT fait les calculs, qui sont donc extrêmement précis puisqu'il est tenu compte de la durée des segments pour la moyenne et l'écart-type. A partir du script écrit pour la F0, nous avons écrit un script pour l'intensité (*measure intensity*)⁸⁶, qui prend les mesures pour toutes les UT et UI d'un segment.

3.3.4 Résumé

Ne pouvant procéder à une analyse de la balance spectrale, qui aurait certainement été plus probante et surtout plus novatrice, nous avons souhaité analyser l'intensité globale en relation avec l'emphase et afin de ne pas trop succomber à la variabilité de l'intensité due à des phénomènes extérieurs, nous avons décidé de rentrer dans notre tableau de données six paramètres pour chaque unité tonale et chaque unité intonative : l'intensité minimale (dB), le temps de l'intensité minimale (s.), l'intensité maximale (dB), le temps

⁸⁶ Le script se trouve en annexe 3b ou sur le CD ROM : *Scripts : PRAAT : measure intensity*.

de celle-ci (s.), l'intensité moyenne (dB), l'écart-type (dB). Comme pour la F0, nous effectuerons des calculs à partir de ces paramètres que nous inclurons dans notre tableau de données⁸⁷.

Le logiciel statistique, faisant le tri des paramètres les plus pertinents en corrélation avec le degré d'emphase, nous dira si l'intensité est un paramètre révélateur et nous pensons que si c'est le cas nous aurons sûrement une augmentation relative de l'intensité.

⁸⁷ Cf. chapitre 4.

3.4 *Pauses*

Les pauses représentent un phénomène bien connu et universel, puisque quels que soient le locuteur ou la langue, on est obligé de faire des pauses lorsque l'on parle, ne serait-ce que pour reprendre son souffle. Mais ce besoin physiologique est utilisé à des fins stratégiques et l'importance des pauses a été mise plusieurs fois en évidence dans la littérature (cf. chapitre 1) en relation avec l'emphase ou la mise en relief.

Lors de l'étiquetage en mots du corpus, nous avons noté toutes les pauses qui apparaissent et nous nous y sommes intéressée de près. Nous avons vu qu'il semblait exister plusieurs sortes de pauses. Nous avons décidé de déterminer plusieurs catégories de pauses.

3.4.1 Différents types de pauses

Pour Goldman-Eisler (1968), une pause est une interruption du mouvement articulatoire, et elle en distingue trois types :

- lors de transitions articulatoires (quand deux occlusives se suivent par exemple), une interruption est nécessaire pour ajuster la position articulatoire ;
- lorsque l'on hésite, on marque un arrêt ;
- lorsque l'on a besoin de reprendre son souffle, une pause est nécessaire.

En ce qui concerne les pauses dues à l'hésitation, Goldman-Eisler précise qu'il peut s'agir de pauses vides (ou non sonores) ou bien de pauses remplies (ou sonores), elles-mêmes divisées en deux sous-catégories : les hésitations vocales (de type « er » ou « um ») et les allongements vocaliques.

Selon Grosjean (1980) pourtant, les pauses de type allongement syllabique n'apparaissent que dans les langues à syllabation ouverte comme le français, qui diffère en cela de l'anglais. L'anglais étant une langue à syllabation fermée, elle utilisera moins cet allongement syllabique au profit d'une utilisation plus systématique d'hésitations vocales.

Repp (1992) affirme que les auditeurs tendent à percevoir une pause après une syllabe allongée même s'il n'y a pas de trace physique d'un silence.

Ces remarques concernant l'allongement syllabique considéré comme une pause sont très intéressantes mais nous ne prendrons pas ce phénomène en compte pour l'analyse des pauses car nous avons par ailleurs étudié la durée segmentale en rapport avec l'emphase. Il serait donc redondant de considérer ce phénomène dans les pauses.

C'est sur tous les autres types de pauses que nous souhaitons porter notre attention.

Guaitella (1991) distingue deux types de pauses : la pause (interruption vocale), qui est un fait segmental directement mesurable, et l'hésitation vocale (interruption verbale), qui est un fait linguistique. Toutes deux peuvent avoir un rôle de programmation des segments. Qu'en est-il alors de la programmation de l'emphase ? Y-a-il plus de pauses dans un groupe intonatif contenant de l'emphase ?

Grosjean et Deschamps (1972 ; 1975) ont comparé le français et l'anglais et trouvé des similitudes et des disparités intéressantes. Ils constatent que 85% des pauses non sonores en français sont en partie ou en totalité des pauses de prise de souffle, contre 34% seulement en anglais. L'anglais présente plus de pauses mais qui sont moins longues, car il est plus concis.

Cela nous amène à nous demander si la durée des pauses est un phénomène révélateur pour l'emphase. Les pauses entourant des mots emphatiques seraient-elles plus longues ou plus courtes que les autres pauses ?

Pour répondre à toutes ces questions, nous avons d'abord nous aussi classé les pauses en plusieurs catégories, et nous n'avons pas pris en compte les pauses sonores. Ce choix peut paraître étrange au vu des travaux antérieurs qui s'accordent tous à reconnaître l'existence de pauses sonores ou non vides. Pourtant, nous avons regardé de près dans notre corpus les pauses qui se trouvaient juste avant ou juste après les mots donnés emphatiques par les deux tiers des auditeurs : il n'y a aucune occurrence de pause non vide.

3.4.1.1 Pauses et unités intonatives

Les pauses permettent souvent au locuteur de reprendre sa respiration. Ce sont des pauses physiologiques dans certains cas puisqu'on ne peut continuer à parler sans reprendre son souffle. Nous en avons un excellent exemple dans le corpus⁸⁸ :

☞ **exs.40** : P1. 4S07 : PM : *we're going to say well it's going to affect us even if some of these people go in for it it's going to affect us anyway so = were going to be in there*

On entend très bien dans ce segment que PM est obligé de reprendre sa respiration car il a parlé longtemps sans le faire et n'a vraiment plus d'air. C'est clairement à cela que sert la pause.

Ces pauses n'apparaissent pas à n'importe quel moment dans l'énoncé, on choisit logiquement toujours la fin d'un groupe intonatif pour reprendre sa respiration afin de ne pas perturber la communication.

Goldman-Eisler (1968) prétend pourtant que les pauses peuvent intervenir de manière erratique et par conséquent nuire justement à la communication. Nous en avons quelques exemples dans nos segments : dans P1.2S05 (☞ **exs.41**), on a une respiration au milieu du groupe intonatif⁸⁹ :

/ Don't forget / there was also in the press today that about eightyfive to a hundred Tories = are gonna have a different policy on Europe /

Cependant, les pauses ne sont pas perçues de la même façon selon qu'elle apparaissent à des endroits attendus ou non. Dans ce dernier cas, elles sont moins souvent perçues (Repp, 1992). Par conséquent, elles ne gênent pas tant que cela la communication, contrairement à ce que pense Goldman-Eisler.

Nous avons donc mis dans la même catégorie les pauses qui permettent de reprendre son souffle et celles qui séparent tout simplement deux groupes intonatifs, sans prise de souffle⁹⁰. Dans ce dernier cas, la pause n'aura qu'une fonction pragmatique, à savoir montrer aux interlocuteurs que l'on a fini de parler ou encore que l'on passe à autre chose, etc...

⁸⁸ Nous représentons les pauses par le symbole « = ».

⁸⁹ Les barres obliques marquent les frontières d'unités intonatives.

⁹⁰ Il ne nous a pas semblé utile de distinguer deux catégories ici.

Il faut noter que ces pauses sans reprise d'air sont quand même assez rares (nous n'en avons aucun cas dans notre corpus). On reprend en général à chaque fois de l'air même si l'on n'en a pas vraiment besoin.

Nous devons préciser également que la fin d'une unité intonative n'est pas forcément marquée par une pause, loin s'en faut.

3.4.1.2 Pauses et hésitation

Certaines pauses semblent être dues à une hésitation de la part du locuteur, ou encore au fait que le locuteur réfléchit à ce qu'il va dire. Ceci est clair quand on a une marque d'hésitation comme « er » ou « erm » ou une répétition ou une expression du style « you know ». Parfois aucune marque précise ne permet de savoir s'il s'agit d'une hésitation ou d'un arrêt pour réfléchir, ou planifier le reste du discours, c'est plutôt le contexte ou la qualité de la voix qui permet de le déceler.

Par exemple dans notre passage P2.1S03 (↩ exs.42), VB dit :

No last year I stayed with a friend = er in her studio

La pause est clairement due à une hésitation, le « er » le montre bien.

Dans P3.3S05 (↩ exs.43), PS dit :

and secondly = we've = actually seen quite a lot of er er cases...

Nous pensons que la première pause marque l'unité intonative alors que la deuxième pause marque une hésitation. C'est tout le contexte (la présence de « er er » plus loin) et certainement aussi la qualité de la voix (qui sonne hésitante) qui nous fait affirmer cela.

Il faut noter qu'une pause due à une hésitation est souvent l'occasion pour nos locuteurs de reprendre leur souffle, mais pas systématiquement.

Notons également que nous plaçons dans cette catégorie les pauses qui permettent au locuteur de réfléchir, puisque nous avons vu que dans les deux cas la fonction principale est de programmer sa parole.

3.4.1.3 Pauses et emphase

Nous voyons aussi que les pauses se trouvent souvent autour des mots jugés emphatiques par les auditeurs (nous avons choisi pour cette analyse précise des pauses un « seuil d'emphase » à deux tiers, c'est-à-dire que nous estimons que les mots sont emphatiques quand ils sont donnés par deux auditeurs sur trois). Nous avons volontairement employé le mot « autour » car les pauses ne sont pas forcément immédiatement avant ou après le mot emphatique. Elles peuvent être décalées d'un ou deux ou même trois mots. Nous persistons pourtant à dire que ce sont des pauses que nous appellerons « emphatiques » car dans plusieurs cas elles entourent un passage contenant un mot emphatique.

Dans le passage P3.1S08 (↵ **exs.44**) : YB dit

*It also liberates I think = policemen = social workers and teachers = who HAVE felt trapped = that they can't interfere...*⁹¹

Les deux pauses qui entourent les mots « who have felt trapped » marquent l'emphase selon nous. En effet le mot « have » uniquement est donné emphatique par plus de deux tiers des auditeurs (donné par 93,75% des auditeurs), mais « felt » et « trapped » sont aussi donnés emphatiques par 31,25% et 68,75% des auditeurs respectivement. Seul « who » n'est donné emphatique par personne. On peut peut-être donc parler ici d'un groupe de mots emphatique et non pas d'un seul mot. Ce groupe de mots serait « who have felt trapped » tout entier et les pauses entourent en effet ce groupe.

C'est pourquoi nous avons classé les pauses de ce type dans les pauses emphatiques.

De nombreux cas sont beaucoup plus clairs, quand la pause apparaît juste avant ou juste après le mot emphatique. Il arrive que le mot emphatique soit mis en relief grâce à deux pauses, une avant et une après. Voici trois exemples, le premier avec une pause avant le mot emphatique, le deuxième avec une pause après, et le troisième avec deux pauses, avant et après :

↵ **exs.45** : P1.4S04 : PM : *there's a perfectly sensible policy which the Conservative = CABINET the Conservative government...*

⁹¹ Les majuscules marquent le mot emphatique.

☞ **exs.46** : P1.5S01 : DD : *you're one of the ones you're one of the FEW = who are said to ...*

☞ **exs.47** : P3.1S06 : YB : *Well I think the one thing it achieves just by = EXISTING = is telling people...*

3.4.1.4 Ambiguïtés et prise de souffle

Certaines pauses sont difficiles à catégoriser. Il semble qu'elles soient du type emphatique mais correspondent aussi à la fin d'une unité intonative par exemple dans P1.4S03 (☞ **exs.48**), locuteur PM :

/ there's a great deal of creative JOURNALISM = / and this is creative JOURNALISM /

Nous avons également un cas où MM est obligée de reprendre sa respiration, on sent qu'elle est à bout de souffle, et cette pause vient aussi après deux mots emphatiques et avant deux autres mots emphatiques dans P1.8S08 (☞ **exs.49**) :

...what the state of the economy is whether the convergence criteria in relation to unemployment and the things that BOTHER US = are ACTUALLY HAPPENING

Nous avons alors pensé que le critère de la respiration pouvait être pertinent car en écoutant s'il y avait des respirations à chaque pause nous avons remarqué que toutes les pauses qui correspondaient à des fins d'unités intonatives et qui étaient non emphatiques étaient caractérisées par une prise de souffle. Par contre, lorsque la pause nous semble clairement emphatique, il n'y a jamais prise de souffle, sauf chez une locutrice, dans l'enregistrement de la BBC : YB.

Ainsi est-il peut-être possible de désambiguïser certains cas :

☞ **exs.50** : P1.5S07 : PM : *EVERYTHING that we've been looking for = and = that is the judgement of the MAN = and led by the Prime Minister*

Dans cet exemple, la première pause semble être due à la fin de l'unité intonative, PM respire. La deuxième pause ne contient pas de respiration, et n'intervient pas à la fin d'un groupe intonatif. Elle pourrait être due à une hésitation ou à une emphase. Or le

mot « man » n'a pas été donné emphatique par au moins deux tiers des auditeurs. Mais il a quand même été donné par 8 auditeurs sur 18 (44,44%). Ceci nous amène au problème suivant : est-ce qu'une pause due à une hésitation ne serait pas susceptible de faire percevoir le mot qui suit comme emphatique ? Cela pourrait être un critère d'erreur de la part des auditeurs et cela se produit en effet dans la conversation de tous les jours. Un cas comme celui-ci est donc tout à fait ambigu. Dans notre exemple, la troisième pause est un cas qui mérite également d'être examiné : elle correspond à la fin d'une unité intonative et vient juste après une emphase. Dans quelle catégorie va-t-on alors la classer ? C'est dans ce type de cas que la respiration peut aider à trancher. Ici le locuteur ne reprend pas son souffle, il nous semble donc que cette pause est emphatique, sinon le locuteur n'aurait eu aucune raison de s'arrêter.

Nous avons d'autres cas ambigus : dans

☞ **exs.51** : P1.8S03 : MM : *our position is = that a) there ought to be a referendum = so people...* ,

exactement comme dans l'exemple précédent, la deuxième pause ne servant pas à respirer peut être soit une pause pour réfléchir, soit une pause emphatique (« referendum » est donné emphatique par 10 locuteurs sur 18 : 55,55%). Dans

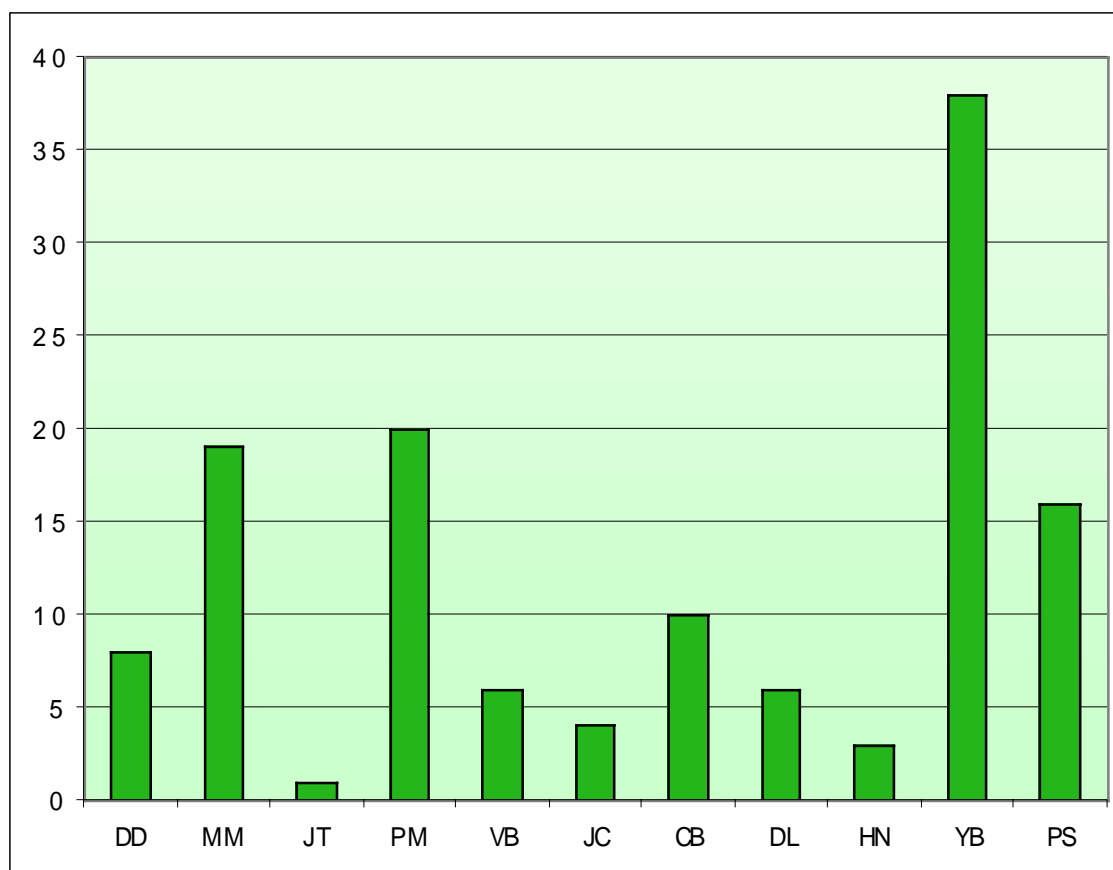
☞ **exs.52** : P1.8S04 : MM : *secondly = in PRINCIPLE = we believe in ...*

la deuxième pause peut être soit emphatique soit respiratoire (MM reprend son souffle). Dans de tels cas, ne pouvant trancher, nous avons classé une même pause dans deux catégories.

3.4.2 Nombre de pauses

Nous avons pensé qu'il serait peut-être intéressant de faire une étude sur le nombre de pauses utilisées par chaque locuteur et en particulier de voir si les locuteurs utilisent en général plus de pauses pour marquer une emphase, pour réfléchir ou pour respirer, etc...

Grâce à Excel, nous avons pu faire des statistiques et sortir des graphiques, plus aisément lisibles que des chiffres⁹². Tous les locuteurs ne prononçant pas le même nombre de pauses, il nous a semblé plus judicieux de procéder avec des pourcentages car en effet le nombre de pauses est très variable comme le montre le graphique suivant :



Graphique 3⁹³ : nombre de pauses par locuteur

Au vu de ce déséquilibre, nous avons écarté les locuteurs JT, JC et HN qui réalisent moins de 5 pauses (1, 4 et 3 respectivement).

Deux facettes du problème se révèlent intéressantes :

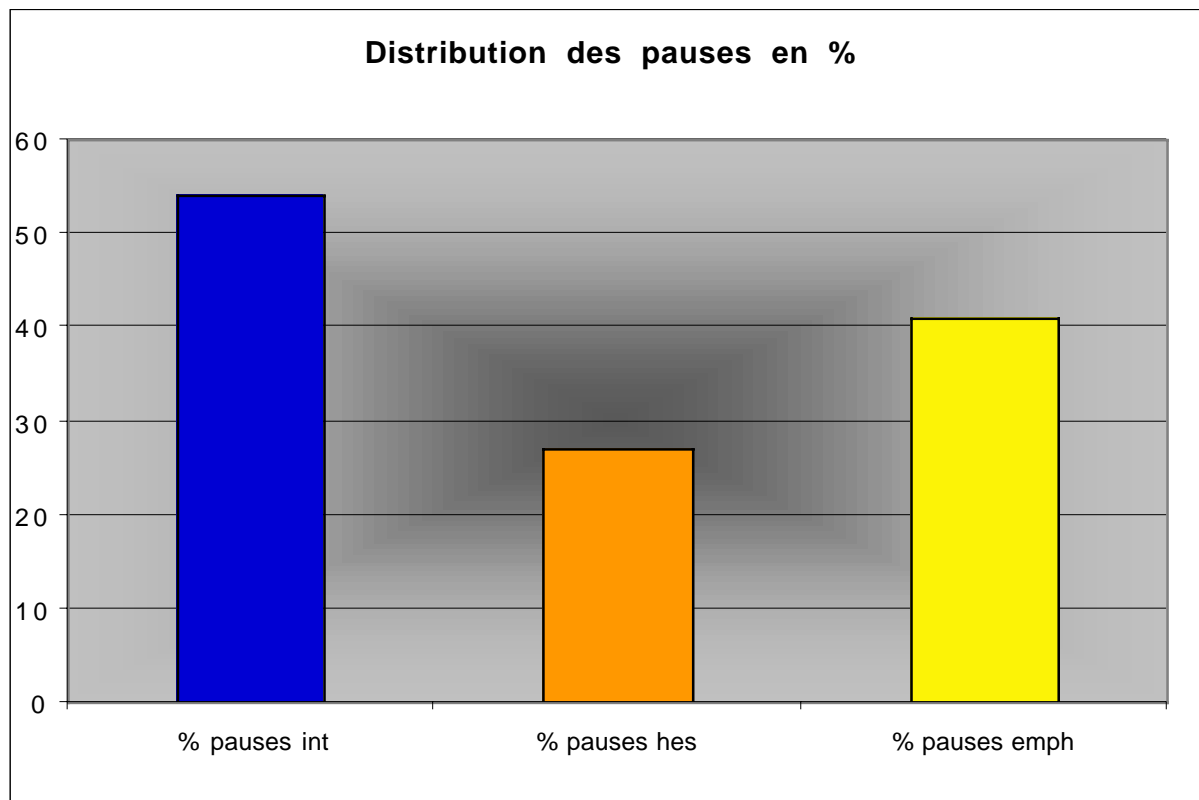
- la distribution entre les pauses des trois catégories que nous avons définies ci-avant, à savoir les pauses qui délimitent des groupes intonatifs (que nous nommons '*pauses int*' et que nous représentons en bleu), les pauses dues à l'hésitation ('*pauses hes*' en orange) et les pauses emphatiques ('*pauses emph*' en jaune) ;

⁹² Cf. CD ROM : *Pauses* ou annexe 3c (les chiffres uniquement).

⁹³ Cf. annexe 3c.

- et la différence entre les locuteurs qui emploient peut-être des stratégies diverses ou ont des habitudes propres.

La distribution des pauses selon notre classement (il faut noter que certaines pauses peuvent avoir été classées dans deux catégories, ‘emph’ et ‘int’ par exemple ou ‘emph’ et ‘hes’ lorsque nous estimions que ça pouvait être l’un ou l’autre, cf. 3.4.1.4), tous locuteurs confondus (à l’exception de JT, JC et HN), est donc la suivante :



Graphique 4⁹⁴ : distribution des pauses avec des pourcentages en abscisses

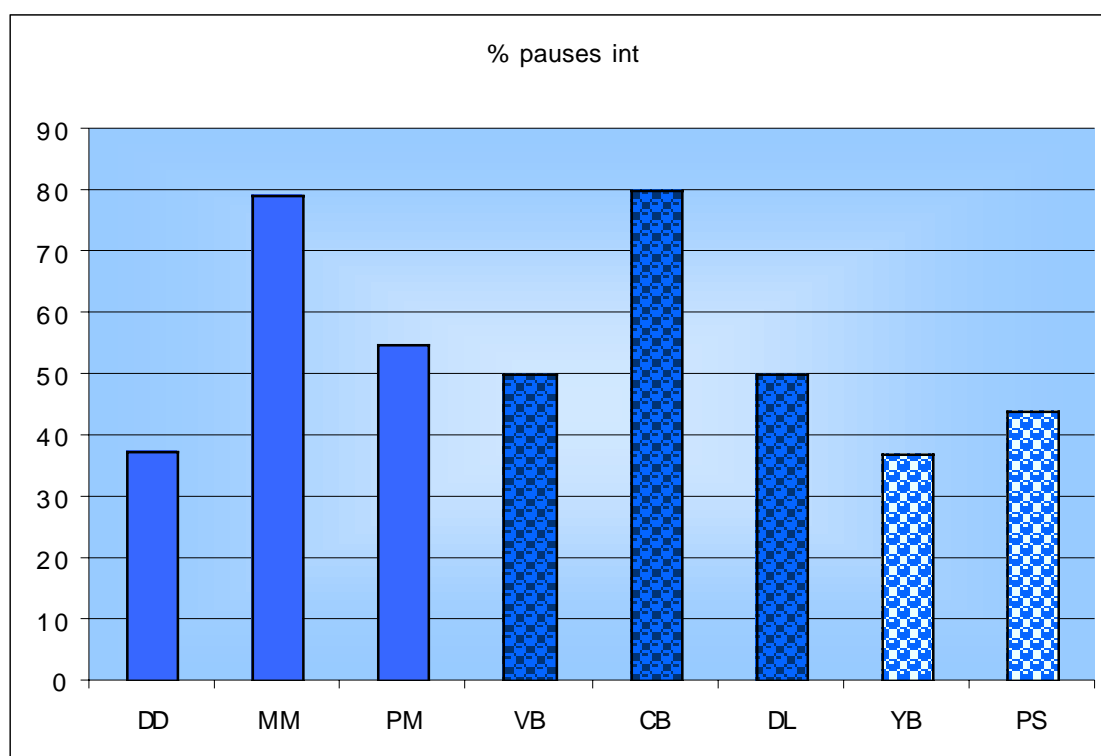
Le fait d’avoir un plus grand nombre de ‘*pauses int*’, c’est-à-dire qui correspondent à des fins d’unités intonatives, n’est pas étonnant. C’est la fonction principale de la pause. On note qu’il n’y a pas énormément de pauses dues à l’hésitation, on est pourtant en discours spontané et on pourrait penser que les locuteurs planifient moins le discours à l’avance. Bien sûr il est très intéressant de voir que l’on a plus de 40% des pauses qui sont classées emphatiques, qui entourent donc un mot donné emphatique. Cela nous encourage à continuer notre étude sur les pauses qui semblent donc avoir une relation avec l’emphase.

⁹⁴ Cf. annexe 3c.

Nous avons ensuite observé le comportement de chaque locuteur et avons pensé qu'il ressortirait peut-être des nuances par locuteur mais aussi en fonction du type de discours. C'est pourquoi chaque type de discours est différencié sur les histogrammes des graphiques 5, 6 et 7⁹⁵ :

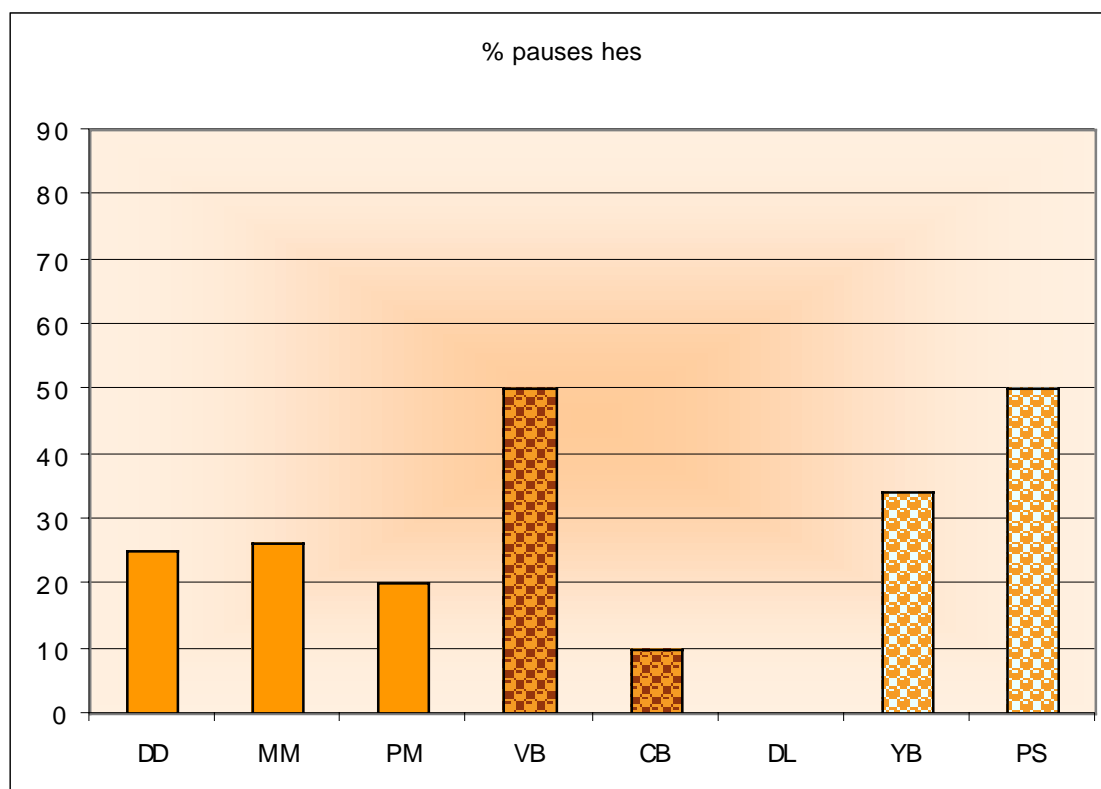
- les trois premiers locuteurs parlent dans le débat politique, en uni sur chaque histogramme ;
- les trois locuteurs suivants sont ceux de la conversation, représentés par les barres avec des sphères foncées ;
- les deux derniers sont les femmes de l'émission radio, représentées par les barres avec des sphères claires.

Nous avons laissé les mêmes couleurs pour les trois types de pauses :

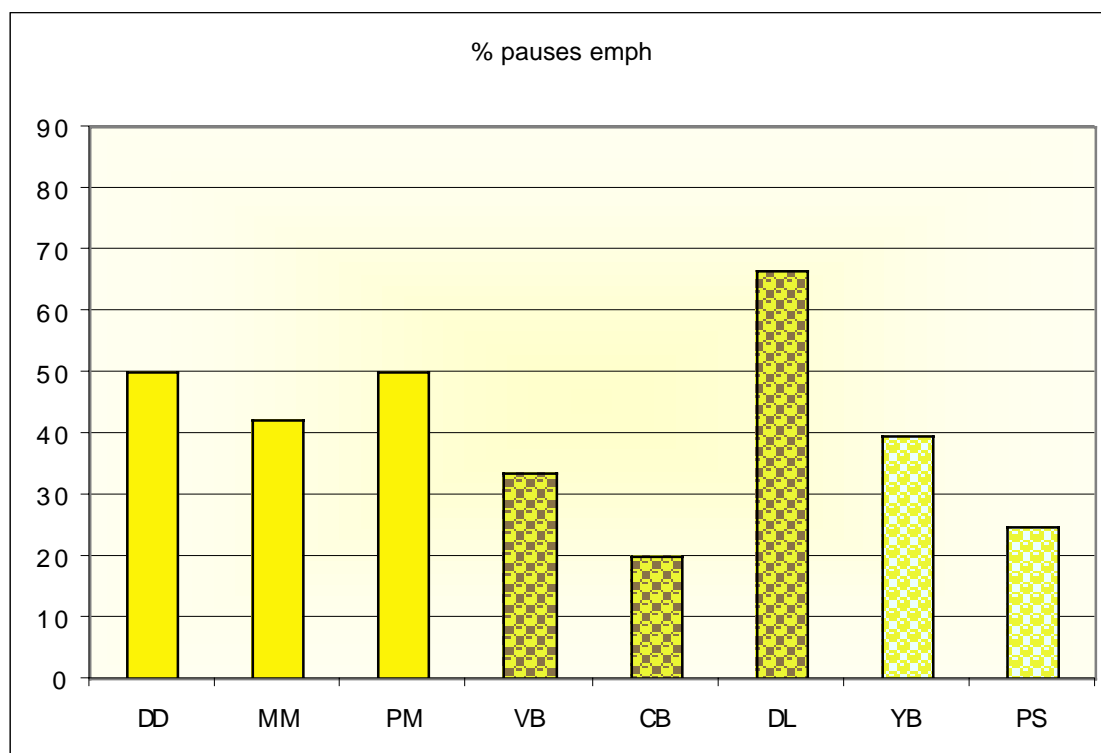


Graphique 5 : pourcentage de pauses marquant des frontières d'unités intonatives par locuteur

⁹⁵ Cf. annexe 3c pour les chiffres correspondant aux graphiques.



Graphique 6 : pourcentage de pauses marquant des hésitations par locuteur



Graphique 7: pourcentage de pauses emphatiques par locuteur

Ces graphiques nous semblent tout à fait révélateurs en ce qui concerne l'emphase et le type de discours également. On remarque tout de suite que le pourcentage de pauses emphatiques varie selon les locuteurs. DL y a plus souvent recours que CB par exemple, qui utilise peut-être plus volontiers un autre procédé. C'est ce que l'analyse statistique nous permettra peut-être de dire.

Quant au type de discours, on peut déjà tirer quelques conclusions. Les trois premiers locuteurs sont des personnages politiques et donc des orateurs. Ils ont l'habitude de s'exprimer en public et savent de quoi ils parlent. Même s'il s'agit de parole spontanée, leur discours est plus ou moins préparé, et on remarque que ce sont eux qui font le plus souvent appel aux pauses emphatiques et le moins souvent aux pauses dues à l'hésitation (ou planification). Le faible nombre de pauses dues à l'hésitation s'explique probablement aussi par le type de discours qui est un débat et si une personne hésite trop, elle se fait couper la parole, alors que ce n'est pas le cas pour la conversation de la BBC (YB et PS, sphères claires) où les locutrices ne sont pas dans le même type d'interaction et ont bien plus le temps d'hésiter et de réfléchir. PS n'a d'ailleurs que très peu recours aux pauses emphatiques mais cherche souvent ses mots. Elle est en effet au cœur du sujet, ayant été elle-même confrontée au type de problème dont elle parle. Son discours est donc bien plus émotionnel et c'est sûrement pourquoi elle hésite autant. Les pauses sont réparties de façon équilibrée chez YB qui est plus virulante et a un discours plus politique que son interlocutrice car elle représente une association. Cependant le sujet est délicat et la touche beaucoup aussi, d'où la différence avec le discours purement politique.

En ce qui concerne les locuteurs de la conversation de tous les jours (sphères foncées), ils ne sont jamais en désaccord et ne cherchent pas à se convaincre. Il est donc intéressant de voir si leurs techniques diffèrent de celles des cinq autres. On remarque tout de suite que comme PS, VB hésite et cherche plus souvent ses mots. Ceci n'est cependant pas valable pour DL ni CB. DL a un débit de parole assez rapide comparé aux autres, ceci expliquant peut-être cela. Quant aux pauses emphatiques, VB et DL y ont assez souvent recours, mais CB les utilise très peu. CB a pourtant, selon les auditeurs, recours à l'emphase comme les autres.

Il découle de toutes ces remarques que les pauses sont très certainement un moyen d'exprimer l'emphase ou en tout cas qu'elles aident à le faire : les pauses emphatiques représentent pour les 8 locuteurs que nous avons choisis 40% du nombre total de pauses.

Nous pensons pouvoir aussi affirmer que les pauses constituent un choix de la part du locuteur, que ce dernier peut y avoir recours ou non. Ce n'est pas un procédé obligatoire. CB par exemple emploie certainement d'autres procédés : les chutes sont peut-être plus amples chez cette locutrice. Nous avons vu qu'elle part parfois d'assez haut dans l'aigu et cela est un indice d'emphase chez elle. L'analyse statistique devrait nous permettre de vérifier cette hypothèse si nous l'appliquons aux données de chaque locuteur.

Cette analyse préliminaire sur le nombre de pauses nous conforte encore une fois dans l'idée qu'il s'agit d'un paramètre non négligeable en relation avec l'emphase, et nous l'intégrerons donc dans les données statistiques.

3.4.3 Durée des pauses

Goldman-Eisler (1968) a fixé la durée minimale d'une pause silencieuse à 250 ms. En deçà, cette durée peut correspondre à la période de tenue d'une occlusive. En général, les pauses ne dépassent pas 3 secondes et 99% des pauses sont même inférieures à 2 secondes. Les pauses respiratoires vont de 0,5 à 1 seconde.

Le logiciel PRAAT nous permet d'avoir la durée des pauses. Nous pouvons donc voir s'il y a une différence de durée entre les divers types de pauses déterminés ci-dessus, et surtout de voir si les pauses de type emphatique se distinguent des autres ou non.

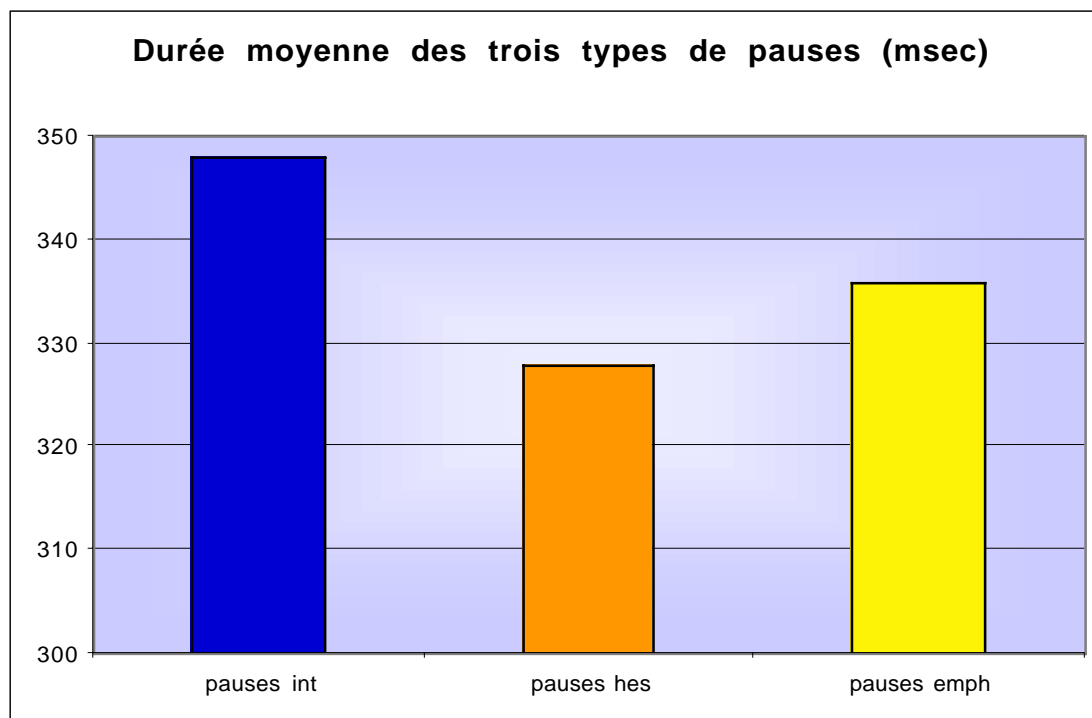
Nous avons procédé comme pour le nombre de pauses, en rentrant nos données dans une feuille de calcul Excel et en sortant des graphiques⁹⁶. Comme ci-dessus, nous avons écarté JT, JC et HN.

Nous avons calculé la durée moyenne selon le type de pauses :

- pauses int : 348 msec ;
- pauses hes : 327,8 msec ;

⁹⁶ Cf. CD ROM : *Pauses* ou annexe 3d pour les chiffres totaux uniquement.

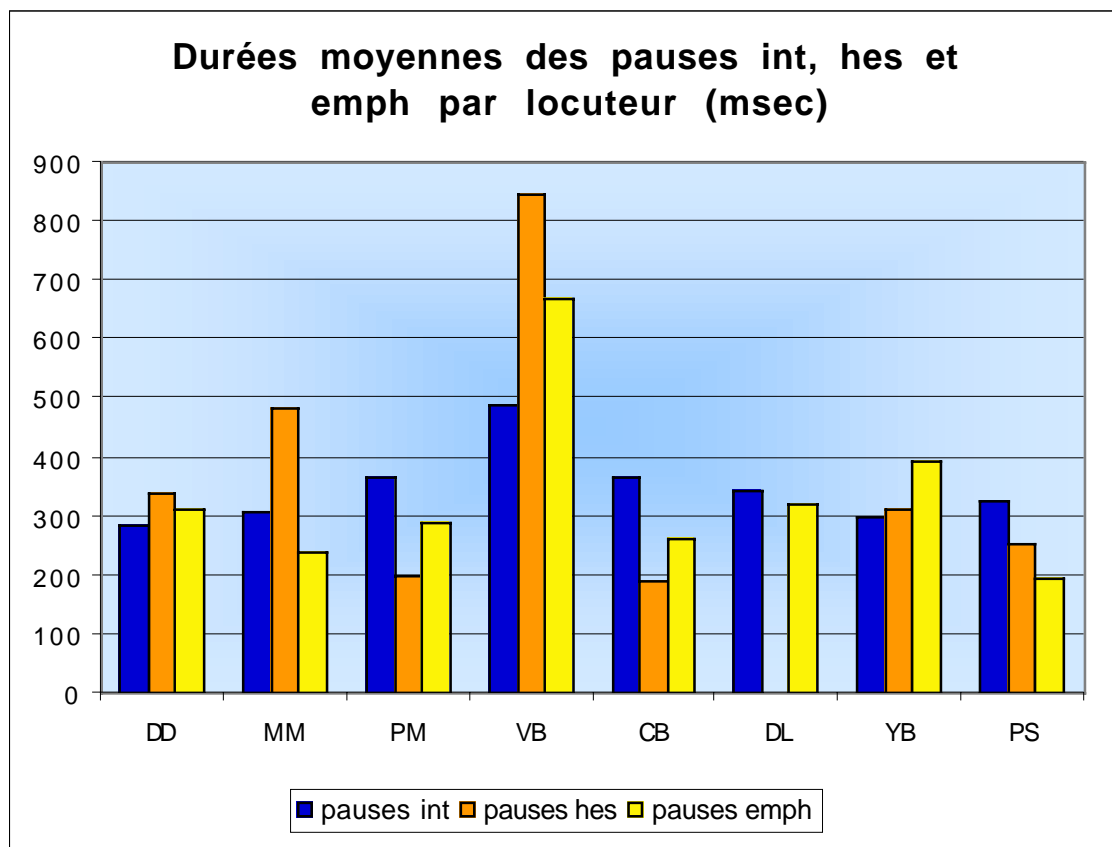
- pauses emph : 335,7 msec :



Graphique 8 : durées moyennes en millisecondes des trois types de pauses

Nous remarquons que les pauses dues à l'hésitation sont les moins longues. Ceci peut paraître étonnant, mais n'oublions pas que nous n'avons pris en compte que les pauses silencieuses. Lors d'hésitations, les locuteurs remplissent souvent les pauses.

Nous avons ensuite procédé par locuteur, pour voir si tous se comportaient à peu près de la même façon en ce qui concerne les durées :



Graphique 9⁹⁷ : durées moyennes en millisecondes des trois types de pauses par locuteur

Ici, le graphique ne semble pas révélateur. Les pauses emphatiques sont en moyenne plus longues chez YB, mais plus courtes chez les autres locuteurs. Les différences ne nous semblent pas très significatives, comme nous l'avions vu sur le graphique précédent avec les durées pour tous les locuteurs confondus. On ne peut donc tirer aucune conclusion en ce qui concerne la durée des pauses ici. Il sera cependant peut-être intéressant de voir s'il existe une corrélation entre le degré d'emphase et la durée de la pause qui est proche du mot emphatique.

3.4.4 Pauses emphatiques : avant ou après ?

La dernière question que nous sommes posée à cette étape, à propos des pauses, concerne uniquement les pauses que nous avons appelées emphatiques : apparaissent-

⁹⁷ Cf. annexe 3d.

elles plus souvent avant ou après le mot emphatique ? Peut-on en tirer un principe général ?

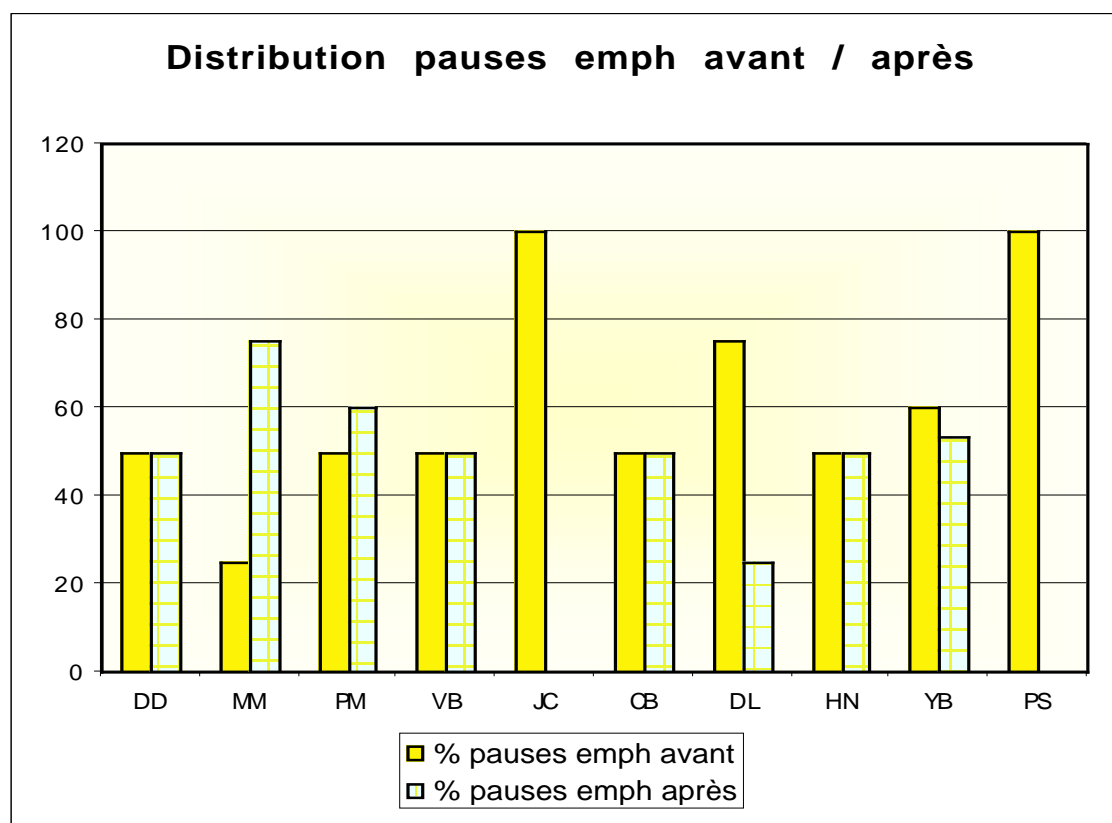
Nous avons classé les pauses emphatiques en « pauses emph avant » et « pauses emph après ». Nous n'avons cette fois éliminé que JT car il ne fait qu'une pause emphatique et avons gardé tous les autres locuteurs. Nous avons été confrontée deux fois à une pause qui se trouvait entre deux mots emphatiques. Nous avons mis ces pauses dans les deux catégories. Les chiffres sont les suivants (il s'agit du nombre de « pauses emphatiques », de « pauses emphatiques avant » et de « pauses emphatiques après ») :

locuteurs	Nbre de pauses emph	Nbre de pauses emph avant	Nbre de pauses emph après
DD	4	2	2
MM	8	2	6
PM	10	5	6
VB	2	1	1
JC	2	2	0
CB	2	1	1
DL	4	3	1
HN	2	1	1
YB	15	9	8
PS	4	4	0

Tableau 1 : nombre de pauses emphatiques, de pauses précédant un mot emphatique, de pauses suivant un mot emphatique

Le graphique 10 ci-dessous montre les résultats de ce classement en pourcentage. Ici encore, on peut difficilement tirer des conclusions claires. On peut tout au plus parler de stratégie propre à chaque locuteur. Comparons MM et JC : la stratégie est inverse. Il est vrai néanmoins que JC ne réalise que 2 pauses (cf. tableau 1) et que l'on peut arguer que cela n'est pas assez représentatif.

Si l'on prend MM, PM et YB, qui sont ceux qui réalisent le plus de pauses emphatiques (cf. tableau 1), on voit qu'il est impossible de dégager une tendance commune.



Graphique 10⁹⁸ : parmi les pauses emphatiques, pourcentage de celles qui se trouvent avant et de celles qui se trouvent après un mot emphatique, par locuteur

3.4.5 Résumé et conclusion

Grâce aux tests de perception réalisés au début de ce travail, nous avons pu extraire des mots donnés emphatiques par la plupart des locuteurs. Cela nous a permis de tester empiriquement les affirmations que l'on trouve dans la littérature relatant l'importance du rôle des pauses dans la perception de l'emphase. Nos résultats montrent que 35,52% des pauses vides de notre corpus primaire sont des pauses emphatiques.

Les différentes études que nous avons faites sur les pauses (nombre, durée, position) ne se sont pas révélées probantes. Nous n'avons pas réussi à dégager un paramètre commun à tous les locuteurs.

⁹⁸ Cf. annexe 3e.

Nous pouvons par contre affirmer que les locuteurs peuvent avoir des stratégies différentes pour ce qui a trait aux pauses (et cela corrobore les affirmations de plusieurs chercheurs, pas uniquement à propos des pauses d'ailleurs, cf. 1.2.3.3). Certains vont faire davantage usage des pauses pour aider à marquer l'emphase et vont placer ces pauses avant le mot ou le groupe emphatique. D'autres vont les placer après. D'autres encore n'auront que très peu recours aux pauses.

Ces stratégies différentes pour un paramètre donné nous poussent à faire une étude par locuteur et à comparer les résultats trouvés pour chaque paramètre. Si CB n'a pas recours aux pauses pour marquer l'emphase, utilise-t-elle davantage un autre paramètre ? Si on la compare à DD par exemple, qui proportionnellement utilise beaucoup plus les pauses emphatiques, va-t-on trouver un rééquilibrage ? DD utilise-t-il moins un autre paramètre que CB ?

C'est ce que nous allons essayer de voir dans le chapitre 4, grâce à une analyse statistique poussée.

Mais auparavant, nous souhaitons traiter de quelques cas particuliers.

3.5 Cas particuliers

Tout au long de ce chapitre, nous avons cherché à déterminer quels paramètres pourraient être pertinents quant à la perception de l'emphase. Nous avons parlé de mouvement mélodique plus ample, qui part souvent d'une fréquence plus élevée que pour le reste de l'unité intonative, ou encore de hausse de l'intensité, ou bien d'un allongement des durées. Toutes ces remarques sont valables d'un point de vue global et général.

Nous souhaitons cependant nous pencher sur des cas particuliers, qui sont assez isolés, ou bien difficiles à mesurer de façon automatique, mais qui nous semblent pertinents dans le cadre de l'emphase. Ce sont des paramètres que nous ne pourrions pas intégrer à notre étude statistique ou qui sont si isolés qu'ils ne ressortiront pas. C'est pourquoi nous les traitons ici.

3.5.1 Baisse de l'intensité

Nous avons vu lors de nos pré-analyses sur l'intensité globale que les mots emphatiques sont souvent parmi ceux dont l'intensité est élevée.

Nous pensons que le phénomène inverse, c'est-à-dire une baisse de l'intensité, peut être un marqueur d'emphase également dans certains cas. Nous en avons un exemple tout à fait révélateur dans notre corpus. VB prononce une phrase dans laquelle elle met clairement de l'emphase :

P2.5S03 (☞ **exs.53**) : *Yeah I had the Duchess of Kent*

La figure 28 ci-dessous présente le fichier son accompagné de son TextGrid où l'on peut voir le pourcentage d'emphase pour chaque mot, et la courbe du bas est la courbe d'intensité (« intensity tier »). On voit nettement sur la courbe d'intensité qu'il y a une baisse. Le mot le plus emphatique (« Kent », en bleu sur la figure 28) est aussi celui dont l'intensité est la plus basse :

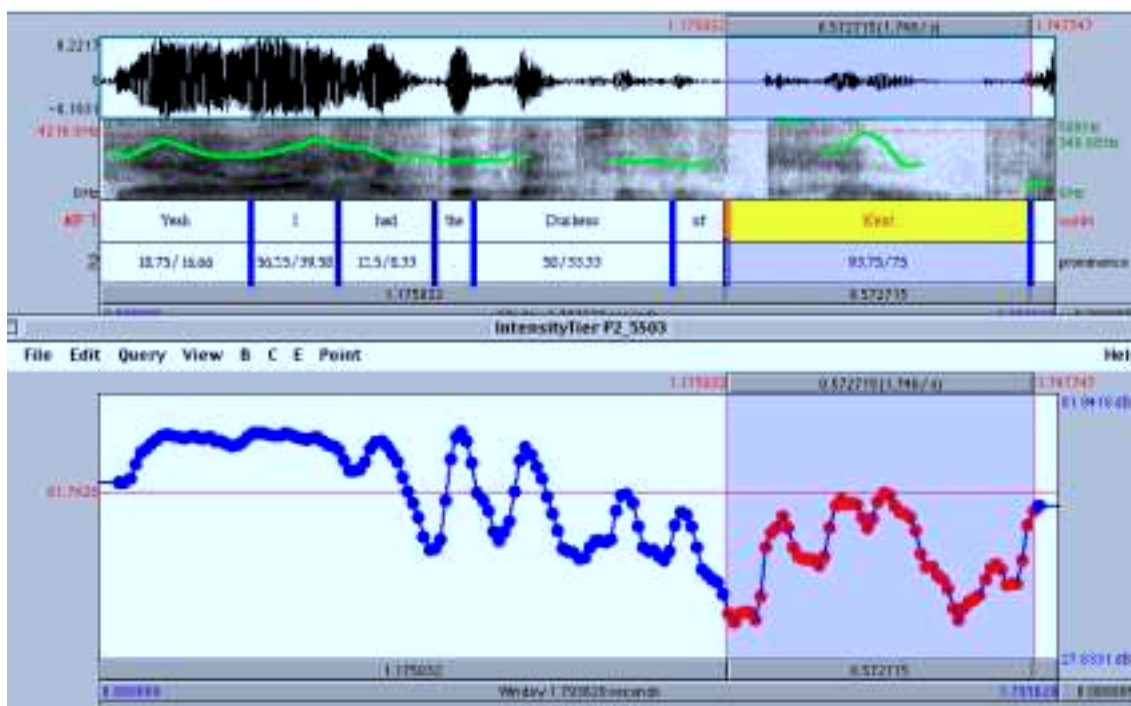


Figure 28 : P2.5S03, son , TextGrid et « intensity tier ».

Il ne s'agit aucunement ici d'une interférence quelconque. Le contexte en est garant. Il s'agit au contraire clairement d'une stratégie oratoire de la part de la locutrice qui met ce mot en valeur en le prononçant moins fort, pour le mettre d'autant plus en relief.

Nous ne disposons dans tout notre corpus primaire que d'un seul exemple de ce type. C'est pourquoi nous le classons dans les cas particuliers. Nous sommes cependant persuadée que ce procédé n'est pas rare ou exceptionnel mais relève d'une stratégie consciente de la part du locuteur. La locutrice VB est enseignante et use peut-être de procédés plus divers que quelqu'un dont le métier ne le pousse pas à attirer et garder l'attention d'un auditoire.

3.5.2 Attaques des mots et coups de glotte

La façon dont le locuteur prononce le début des mots semble déterminante dans la perception de l'emphase :

Quand l'accent contrastif occupe la place de l'ancien accent primaire, il se signale soit par une chute plus importante, soit par une tenue prolongée de la

consonne initiale, ou devant voyelle par l'introduction d'un « glottal stop ».
(Lilly & Viel, 1999 : 187)

En ce qui concerne la chute, les paramètres que nous mesurons peuvent en rendre compte. Ce n'est par contre pas le cas pour ce qui est de la tenue prolongée de la consonne initiale puisque nous mesurons les durées sur les mots, et non sur les phonèmes. Quant aux coups de glotte, il nous est impossible de les incorporer dans une analyse statistique. C'est pourquoi nous traitons ces points dans les cas particuliers, alors que ce n'en sont pas.

Ce que l'on appelle traditionnellement l'attaque d'une syllabe (par opposition à la rime) est la consonne ou le groupe consonantique qui se trouve à l'initiale de la syllabe. Nous nous intéressons pour notre part à la syllabe initiale du mot. Il semble que l'un des moyens de mettre un mot en relief est donc d'allonger la (ou les) consonne(s) initiale(s). Certains exemples illustrent très bien ce phénomène :

P1.1S03 (↵ exs.54) : *the scooters his kid's scooters*

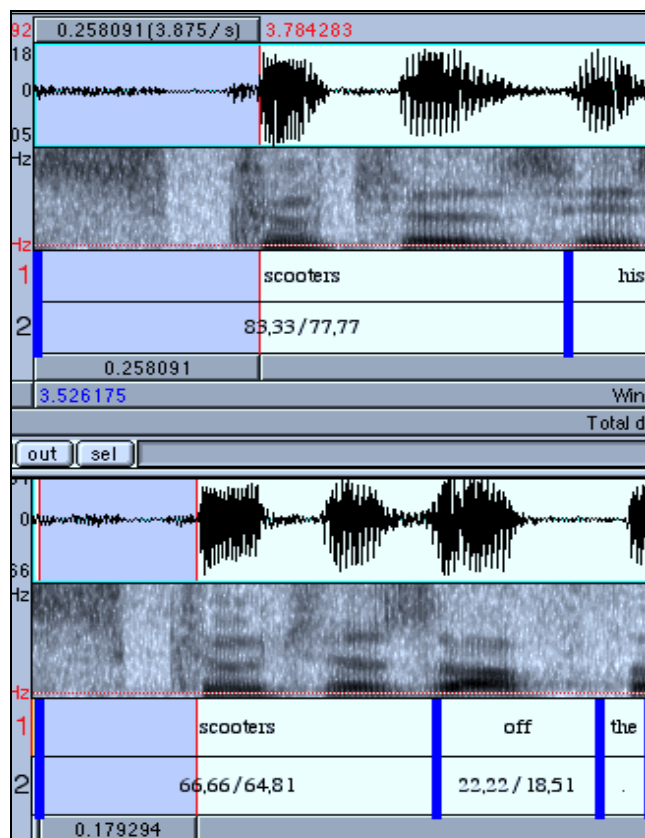


Figure 29 : P1.3S03 : longueur du groupe consonantique initial de « scooters »

Dans cet exemple, les deux occurrences de « scooters » sont emphatiques et sur la figure 29 le groupe consonantique /sk/ se trouve dans le bleu. Le chiffre sous la sélection en bleu est la durée du groupe en secondes : on voit que /sk/ a une durée de 0,258 sec. dans le premier cas et de 0,179 sec. dans la deuxième occurrence. Cette dernière est d'ailleurs moins emphatique. D'après les données de Campbell (cf. 3.2.1.2), la durée moyenne d'un /s/ est de 0,116 sec. (avec un écart-type de 0,051 sec.) et celle d'un /k/ de 0,081 sec. (écart-type : 0,033 sec.), ce qui donne 0,149 sec. Dans la première occurrence de « scooters », le groupe consonantique initial est donc fortement allongé (il l'est moins dans la deuxième occurrence). Nous pensons que cet allongement contribue à rendre le mot emphatique.

Un autre exemple avec une plosive initiale illustre à nouveau ce point :

P1.2S05 (↵ **exs.55**) : *a different policy*

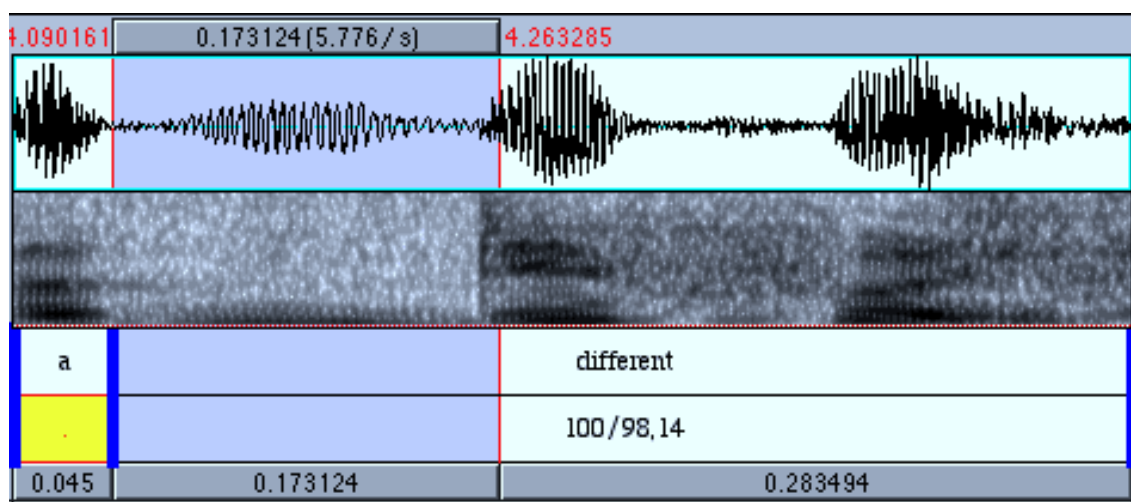


Figure 30 : P1.2S05 : tenue du phonème initial de « different »

La tenue du /d/ est très longue. Nous avons sélectionné en bleu tout le /d/ et voyons qu'il dure 0,173 sec., ce qui est beaucoup plus long que la durée moyenne qui se situe autour de 0,052 sec. (avec un écart-type de 0,034 sec. seulement). Nous ne prétendons pas que la tenue de la consonne initiale est le seul paramètre à rentrer dans la perception de l'emphase. Le mouvement mélodique est certainement aussi pour beaucoup. Mais nous pensons que l'allongement de la consonne initiale renforce l'emphase.

Lilly & Viel mentionnent l'apparition d'un coup de glotte dans le cas d'une voyelle initiale.

Le segment P1.8S08 exemplifie cette affirmation. Il est difficile de voir un coup de glotte sur le spectrogramme, bien sûr, puisqu'il est caractérisé par un blanc, ou sur l'oscillogramme qui sera très plat. On ne peut pas savoir s'il s'agit d'une pause ou d'un coup de glotte. Nous le montrons néanmoins sur la figure 31 ci-dessous (en bleu et symbolisé par le point d'interrogation entre barres obliques), car on peut voir que la durée est de 0,148 sec., ce qui paraît peu pour une pause (cf. 3.4.3).

P1.8S08 (☞ **exs.56**) : *are actually happening*

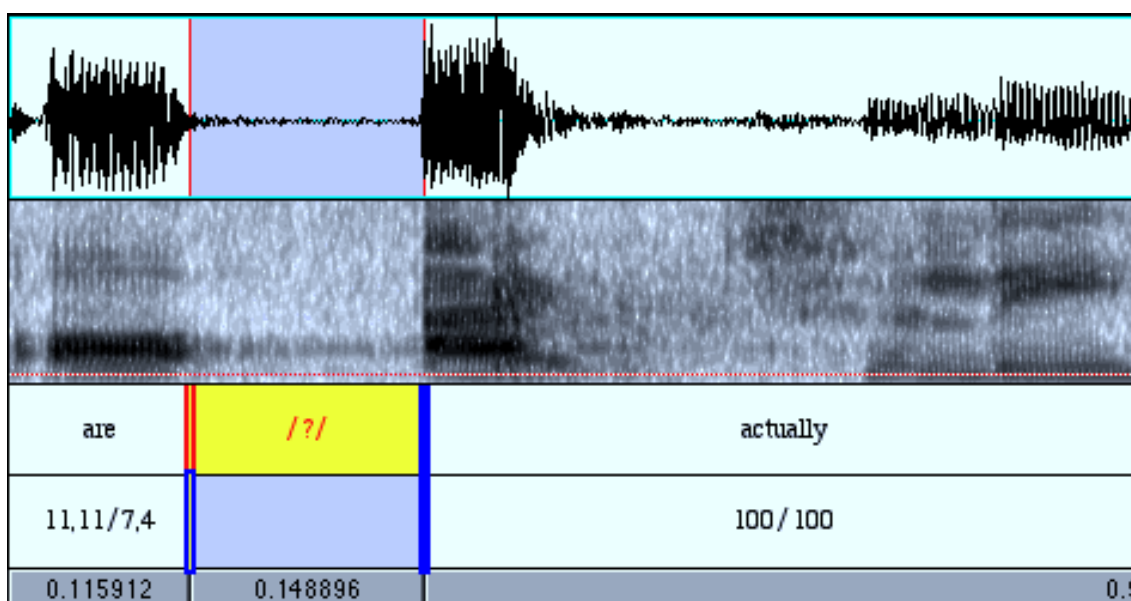


Figure 31 : P1.8S08 : attaque de « actually »

Ainsi, s'il ne s'agit pas d'une réelle pause, le locuteur n'a aucune raison de ne pas faire la liaison entre « are » et « actually ». Il semble en outre que l'on entende bien un coup de glotte à l'écoute. Encore une fois ce coup de glotte n'est pas le seul marqueur d'emphase mais y contribue fortement.

Le même phénomène, encore plus flagrant, se produit dans

P3.1S03 (☞ **exs.57**) : *or actual abuse*

où la durée du silence (0,094 sec.) est encore plus courte que dans l'exemple ci-dessus, ce qui est encore moins équivoque. Il ne peut absolument pas s'agir d'une pause :

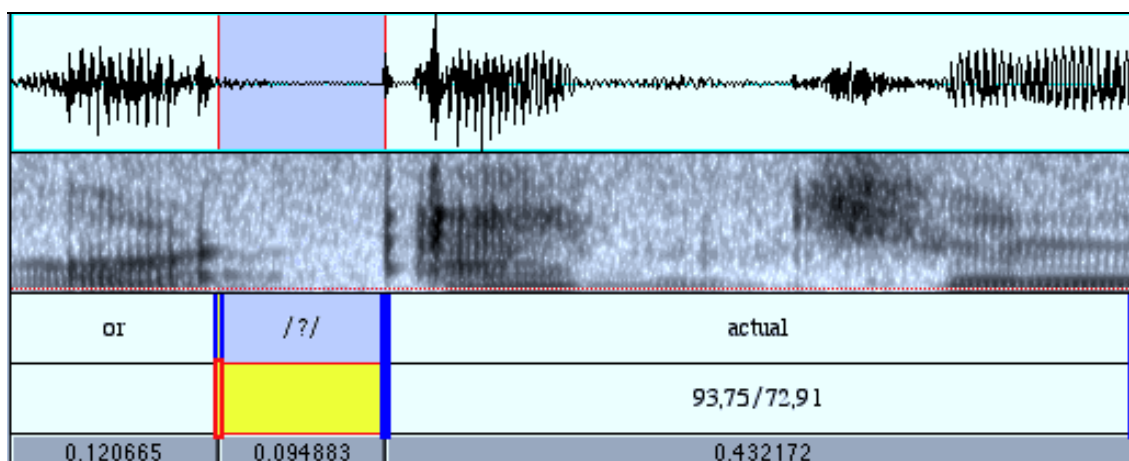


Figure 32 : P3.1S03 : coup de glotte devant « actual »

3.5.3 Qualité de la voix

Le paramètre « qualité de la voix » est souvent mentionné dans les études que nous avons décrites dans le chapitre 1, mais très peu de chercheurs se sont penchés sur cette question. Ní Chasaide & Gobl (2000) ont étudié le phénomène, mais très précisément en relation avec les émotions. Nous pensons qu'il s'agit là d'un facteur qui peut jouer un rôle dans l'emphase. Nous n'en avons qu'un seul cas sans équivoque dans notre corpus primaire :

P2.4S03 (↵ **exs.22**) : *so I thought oh*

Il est impossible d'intégrer ce facteur dans nos analyses et nous pensons que la qualité de la voix n'est pas un des facteurs essentiels de l'emphase, mais mérite d'être mentionnée.

3.6 *Résumé et conclusion*

Cette analyse des paramètres acoustiques et prosodiques de l'emphase nous a permis de vérifier nos hypothèses. Les quatre paramètres que nous avons dégagé des études antérieures semblent en effet révélateurs de l'emphase. De quelle façon et à quel niveau ? C'est ce que ce chapitre nous a laissé entrevoir. Grâce à ces analyses, nous sommes en mesure d'affirmer que les mots emphatiques sont plus souvent porteurs d'un mouvement mélodique ample, qui part de plus haut que pour les mots adjacents. Nous pouvons aussi dire que l'allongement des mots semble être une marque d'emphase. Les mots emphatiques sont en général prononcés de façon plus intense que les autres mots de l'unité intonative. Seules les pauses ne sont pas aussi révélatrices de l'emphase que les trois autres paramètres.

Ces paramètres nous paraissent tout de même difficiles à séparer. Il semble qu'ils soient souvent associés. Une analyse statistique se révèle du fait nécessaire afin de déterminer de façon scientifique quel est le paramètre le plus pertinent. Ceci fait l'objet du chapitre suivant.

Nous avons néanmoins souhaité consacrer une sous-section à des cas particuliers car nous pensons que l'emphase peut être exprimée de maintes façons. L'analyse statistique extraira pour nous les paramètres récurrents, mais nous ne devons pas oublier que ce ne sont pas les seuls moyens de mettre l'emphase sur un mot ou une partie d'un énoncé, les cas décrits en 3.5 en étant la preuve.

4 Chapitre 4

Analyse statistique

Grâce aux travaux décrits dans les chapitres précédents, nous avons dégagé des éléments caractéristiques dont nous allons entreprendre ci-après l'analyse statistique.

En effet, le chapitre 1 nous a permis de déterminer grâce aux études antérieures quels sont les paramètres acoustiques et prosodiques les plus certainement corrélés à l'emphase, à savoir la fréquence fondamentale, la durée, l'intensité et les pauses. Le chapitre 2 décrit tout le travail de création et de préparation du corpus que nous utilisons. Il décrit également la méthode que nous avons appliquée pour déterminer les mots ou passages emphatiques. Le chapitre 3 s'intéresse en détail aux paramètres mentionnés ci-dessus et relate les prises de position et les mesures effectuées. Afin de vérifier nos hypothèses, nous avons donc choisi d'effectuer une analyse statistique. Cette analyse doit nous permettre de déterminer de façon tout à fait scientifique quels paramètres, parmi tous ceux que nous avons extraits, sont les mieux corrélés à l'emphase.

Nous utilisons l'outil statistique CRUISE (Kim & Loh, sous presse ; <http> : cf. références bibliographiques) à cette fin. C'est un des derniers outils de type CART (Classification And Regression Trees), très modulable. Nous aurions pu utiliser un logiciel plus simple, tel que STATVIEW, mais nous avons opté pour CRUISE, plus professionnel et plus performant. Nous sommes consciente qu'il s'agit d'un choix aventureux, ce logiciel ayant été très peu testé dans notre laboratoire avant nous (cf. Bouzon, 2001, pour une utilisation de CRUISE et Bouzon & Blanchard, 2000, pour une utilisation de WAGGON, logiciel de type CART également). Cependant ce choix s'inscrit dans le cadre de notre travail. Nous avons voulu, tout au long de cette étude, bénéficier des outils les plus récents et les plus performants.

Nous décrirons brièvement le logiciel dans la première partie de ce chapitre. Nous expliquerons quels sont les paramètres que nous avons calculés en plus de ceux déjà

donnés dans le chapitre précédent, quels sont les fichiers d'entrée que nous utilisons et les fichiers de sortie auxquels nous parvenons.

La deuxième partie du chapitre sera consacrée à la description des essais préliminaires que nous avons dû mettre en œuvre pour optimiser l'utilisation du logiciel.

Enfin, la troisième partie sera consacrée à l'analyse du corpus, à l'interprétation des résultats obtenus et aux conclusions qui en découlent.

4.1 CRUISE : généralités

Le logiciel de statistique CRUISE (Classification Rule with Unbiased Interaction Selection and Estimation) a été élaboré par H. Kim et W-Y. Loh dans le département de statistique de l'université du Wisconsin. CRUISE est un nouveau programme extrêmement complexe, basé sur plusieurs algorithmes, et qui permet d'obtenir des arbres de régression (pour les détails techniques voir Kim & Loh, sous presse, et pour les conseils d'utilisation voir Kim & Loh, 2001). On donne un fichier d'entrée comportant des paramètres, dont une variable dépendante qui est ce que l'on cherche (l'emphase en l'occurrence). Puis le logiciel opère un tri des paramètres en les classant selon leur pertinence. On obtient un arbre de régression qui permet de prédire les classes déterminées dans la variable dépendante.

Contrairement aux autres analyses statistiques (ANOVA, régression multiple, analyse discriminante, etc ...) qui utilisent toutes les données globalement pour toutes les décisions, les arbres de régression choisissent chaque fois le paramètre qui donne la meilleure classification locale.

CRUISE fonctionne sous Windows et Linux.

4.1.1 Fichiers d'entrée

CRUISE requiert trois types de fichiers d'entrée :

un fichier de données (*.dat* pour « data ») :

Ce fichier contient toutes les données que l'on souhaite analyser. Il doit comporter la variable que l'on veut prédire et les variables de prédiction. Les données doivent être séparées par des virgules, des espaces ou encore des tabulations. De ce fait, si les données contiennent des virgules ou des espaces, elles doivent être entre guillemets. Les données peuvent être numériques ou catégorielles et ne doivent pas dépasser quinze caractères.

un fichier test (*.dat* également)

Il n'est pas indispensable, sauf si l'on souhaite valider les prédictions obtenues à partir du fichier de données. Il doit avoir le même format que ce dernier.

un fichier descriptif (*.dsc* pour « description »)

Ce fichier donne au logiciel les informations suivantes : le nom du fichier de données, le symbole qui signifie « donnée manquante » (car on ne peut pas laisser de case vide dans le fichier de données), le numéro, le nom et le type de toutes les variables. Une variable peut être soit catégorielle (le symbole pour ce type sera 'c'), soit numérique ('n'). On peut également exclure une variable de l'analyse. Elle pourra quand même figurer dans le fichier descriptif, mais suivie du symbole 'x'. Enfin, la variable que l'on cherche à prédire est la variable dépendante ('d'). Une seule variable peut être de type 'd'.

4.1.2 Fonctionnement du programme

On peut choisir un fonctionnement très simple, avec toutes les options par défaut. CRUISE permet également un mode de fonctionnement interactif, où l'on peut choisir soi-même toutes les options. Cette solution est bien plus complexe pour le néophyte en statistiques, mais c'est celle que nous avons choisie car elle offre la possibilité d'adapter le fonctionnement du programme à la spécificité de notre étude. Nous décrivons dans la sous-section 4.3 (essais préliminaires) les raisons de nos choix concernant les options.

4.1.3 Fichiers de sortie

Après traitement, le logiciel CRUISE fournit un fichier de sortie (*.out*) dont les caractéristiques sont les suivantes :

- rappel du fichier descriptif ;
- informations concernant le fichier de données (nombre de variables numériques, catégorielles, nombre de classes, etc...) ;
- rappel des options choisies ;

- résultats de l'analyse : toutes les informations nécessaires à la construction de l'arborescence, en particulier pour chaque noeud, les valeurs des paramètres ;
- matrice de classification sur l'ensemble de la population ;
- matrice de classification rapportée à l'échantillon test.

4.2 Paramètres⁹⁹

4.2.1 Variable dépendante (d)

Le but ultime de notre étude sur l'emphase est de mettre en évidence les paramètres acoustiques et prosodiques qui font qu'une emphase est perçue dans le discours. La variable dépendante est donc l'emphase.

Mais on peut graduer l'emphase et l'exprimer de diverses façons.

Les sondages décrits dans le chapitre 2 nous ont permis de dégager, pour chaque mot du corpus primaire, un pourcentage d'emphase (*%_emph*) et un pourcentage de degré d'emphase (*%_degre_emph*). L'un ou l'autre de ces paramètres peut constituer la variable dépendante puisque nous savons qu'elles sont bien corrélées (cf chapitre 2, graphique 2).

CRUISE détermine à partir de la variable dépendante un nombre de classes.

Compte tenu de la méthode utilisée (cf. chapitre 2, les sondages de perception), l'analyse des données brutes conduit à un trop grand nombre de classes qui ne permet pas la construction d'arborescence. La qualité de l'analyse est donc conditionnée par le choix judicieux du nombre de classes, encore faut-il pour cela déterminer les bons critères à retenir : choix d'une médiane, découpage linéaire ou non, découpage symétrique ou asymétrique par rapport à la médiane ...

Le sujet sera traité en 4.3, voici à titre d'exemples les premières classifications retenues :

répartition symétrique non linéaire pour 6 classes (*coef_emph1*) :

- de 0 à 9% ⇒ classe 0 ;
- de 10 à 29% ⇒ classe 1 ;
- de 30 à 49% ⇒ classe 2 ;
- de 50 à 69% ⇒ classe 3 ;
- de 70 à 89% ⇒ classe 4 ;

⁹⁹ La liste des abréviations de tous les paramètres est donnée en annexe 4a.

- et de 90 à 100% \Rightarrow classe 5.

répartition symétrique non linéaire pour 11 classes (*coef_emph2*) :

- de 0 à 4% \Rightarrow classe 0 ;
- de 4 à 14% \Rightarrow classe 1 ;
- de 15 à 24% \Rightarrow classe 2 ;
- de 25 à 34% \Rightarrow classe 3 ;
- de 35 à 44% \Rightarrow classe 4 ;
- de 45 à 54% \Rightarrow classe 5 ;
- de 55 à 64% \Rightarrow classe 6 ;
- de 65 à 74% \Rightarrow classe 7 ;
- de 75 à 84% \Rightarrow classe 8 ;
- de 85 à 94% \Rightarrow classe 9 ;
- de 95 à 100% \Rightarrow classe10 ;

répartition non symétrique et non linéaire pour 4 classes
(*coef_emph3*) :

- de 0 à 14% \Rightarrow classe 0 ;
- de 15 à 44% \Rightarrow classe 1 ;
- de 45 à 74% \Rightarrow classe 2 ;
- de 75 à 100% \Rightarrow classe 3 ;

4.2.2 Durée

Nous avons quatre paramètres en ce qui concerne la durée¹⁰⁰ :

- *dur_reelle_mot* (durée réelle du mot) : variable numérique donnée en millisecondes ;
- *dur_predite_mot* (durée prédite du mot) : variable numérique donnée en millisecondes ;
- *allong_ms_mot* (allongement du mot en millisecondes) : variable numérique ;
- *allong_%_mot* (allongement du mot en pourcentage) : variable numérique.

¹⁰⁰ Cf 3.2 pour les détails.

Les deux dernières variables se calculent à partir des deux premières. Nous excluons donc les durées pour l'analyse statistique et ne garderons que les allongements. Cependant, il serait redondant de conserver les deux. En conséquence, nous opterons pour l'allongement en pourcentage, qui est significatif puisqu'il exprime un allongement relatif.

4.2.3 Fréquence fondamentale

Le programme PRAAT nous a permis d'obtenir 24 paramètres que nous rappelons brièvement ci-dessous¹⁰¹. Ce sont des variables numériques.

- <i>MinF0_mot</i>	- <i>Mean_UT</i>
- <i>TMinF0_mot</i>	- <i>SD_UT</i>
- <i>MaxF0_mot</i>	- <i>MAS_UT</i>
- <i>TMaxF0_mot</i>	- <i>MASWOJ_UT</i>
- <i>Mean_mot</i>	- <i>MinF0_UI</i>
- <i>SD_mot</i>	- <i>TMinF0_UI</i>
- <i>MAS_mot</i>	- <i>MaxF0_UI</i>
- <i>MASWOJ_mot</i>	- <i>TMaxF0_UI</i>
- <i>MinF0_UT</i>	- <i>Mean_UI</i>
- <i>TMinF0_UT</i>	- <i>SD_UI</i>
- <i>MaxF0_UT</i>	- <i>MAS_UI</i>
- <i>TMaxF0_UT</i>	- <i>MASWOJ_UI</i>

Parmi ces derniers, seuls 12 (que nous considérons significatifs) seront retenus pour notre analyse statistique (moyennes, écarts-type, pentes). Nous ajoutons à cela de nouveaux paramètres calculés par combinaison des 24 paramètres initiaux (les noms sont parfois tronqués car CRUISE n'autorise que 15 caractères) :

- *RF_mot* (« Rise » ou « Fall » sur le mot) : cette variable est catégorielle. Nous avons comparé les valeurs de F0 maximale et minimale, et le temps correspondant à ces valeurs. Si le *MaxF0_mot* précède dans le temps le *MinF0_mot*, alors on a une chute

¹⁰¹ Cf 3.1 pour les détails. Cf. annexe 4a pour les correspondances des abréviations.

(« fall »), si c'est l'inverse, on a une montée (« rise »). Cela en revient à la formule suivante :

si $T_{maxF0_mot} < T_{minF0_mot}$; "Fall" ; si $T_{minF0_mot} < T_{maxF0_mot}$; "Rise".

Nous parlons en 3.1.2.3 du problème du ton creusé et du ton circonflexe et expliquons qu'avec le paramètre « fin de mot » (*fin_mot*), c'est-à-dire le point en secondes où se termine le mot, on peut savoir s'il s'agit en fait d'un ton creusé ou d'un ton circonflexe et non pas seulement d'une chute ou d'une montée. Nous avons testé cela et nous sommes rendu compte qu'en fait, le *Max_F0* ou le *Min_F0* ne correspondent jamais exactement à la fin du mot. On aurait donc toujours un ton creusé ou un ton circonflexe, ce qui n'est pas exact. Les mesures sont trop précises et nous avons donc abandonné cette idée. Nous n'aurons donc que des « rise » ou des « fall », ce qui nous paraît moins erroné que si nous n'avions que des « fall-rise » ou des « rise-fall ».

- *EcMxF0mt_MeanUI* (écart entre *MaxF0_mot* et *Mean_UI*) : nous avons pris la valeur maximale de F0 sur le mot et y avons soustrait la moyenne de la F0 de l'UI. Comme la valeur *MaxF0* est en Hertz et que la valeur moyenne de l'UI est en demi-tons¹⁰², nous avons converti au préalable les demi-tons en Hertz grâce à la formule suivante :

$$F = 100 * 2^{n/12}$$

avec

- **F** la fréquence en Hertz que l'on cherche,
- **n** le nombre de demi-tons à convertir.

La soustraction nous donne donc une valeur en Hertz. Or nous voulons mettre en évidence un écart, et une valeur en Hertz risque de ne pas être suffisamment significative (cf. 3.1.3). En effet, la tessiture étant différente d'un locuteur à l'autre et surtout d'un femme à un homme, que voudrait dire un écart de 20 Hz sans valeur de référence ? Cela est peut-être beaucoup pour un homme comme PM dont la fréquence fondamentale ne dépasse guère 150Hz, et bien peu pour une femme qui monte aisément jusque 450 Hz. C'est pourquoi nous avons choisi d'exprimer un pourcentage plutôt qu'une valeur, afin d'obtenir un écart relatif. Nous avons pris pour base la moyenne de la F0 de l'UI. Nous arrivons donc au calcul suivant :

$$[(MaxF0_mot - Mean_UI) / Mean_UI] * 100$$

¹⁰² Il s'agit de demi-tons tempérés sur une base 100Hz, comme c'est généralement le cas lorsque l'on analyse la parole.

Ainsi cette variable est exprimée en pourcentage, et elle est numérique, comme toutes celles qui vont suivre. Nous avons converti toutes les valeurs des moyennes (*Mean_mot*, *Mean_UT* et *Mean_UI*) de demi-tons en Hertz, et avons à chaque fois exprimé l'écart relatif en pourcentage par rapport à la valeur de l'UI.

- *EcMxF0UT_MeanUI* (écart entre *MaxF0_UT* et *Mean_UI*) : c'est exactement comme au-dessus, mais nous avons voulu comparer la valeur maximale de l'UT cette fois, et non du mot, à la valeur moyenne de l'UI toujours.

- *EcMxF0mot_UI* (écart entre *MaxF0_mot* et *MaxF0_UI*) : nous soustrayons ici la valeur maximale de F0 sur l'UI à la valeur maximale de F0 sur le mot, puis déterminons l'écart relatif, exprimé en pourcentage.

- *EcMxF0UT_UI* (écart entre *MaxF0_UT* et *MaxF0_UI*) : il s'agit de la même chose que ci-dessus mais on compare l'UT et l'UI et non le mot et l'UI.

- *EcMnmt_MeanUI* (écart entre *MinF0_mot* et *MeanF0_UI*) : tout comme nous mesurons la différence entre la valeur maximale de F0 du mot et de l'UT à la moyenne de l'UI, nous mesurons la différence entre la valeur minimale cette fois du mot puis de l'UT (paramètre suivant) et la moyenne de l'UI. Nous avons bénéficié de la conversion de *MeanUI* en Hertz. Ce paramètre est aussi donné en pourcentage, comme les suivants :

- *EcMnUT_MeanUI* (écart entre *MinF0_UT* et *MeanF0_UI*) : cf. ci-dessus.

- *EcMinF0mot_UI* (écart entre *MinF0_mot* et *MinF0_UI*) : comme pour les valeurs maximales, nous regardons s'il y a un gros écart entre les valeurs minimales.

- *EcMinF0UT_UI* (écart entre *MinF0_UT* et *MinF0_UI*) : exactement comme la valeur précédente, mais on compare l'UT et l'UI et non le mot et l'UI.

- *EcMeanF0mot_UI* (écart *MeanF0_mot* et *MeanF0_UI*) : de même nous comparons les valeurs moyennes de F0 du mot et de l'UT, toutes deux exprimées en Hertz, grâce à la conversion ci-dessus.

- *EcMeanF0UT_UI* (écart *MeanF0_UT* et *MeanF0_UI*) : ici encore il s'agit de la même variable qu'au-dessus, mais la différence se situe entre l'UT et l'UI.

- *EcMASmot_UI* (écart entre *MAS_mot* et *MAS_UI*) : on s'intéresse à l'écart entre la pente sur le mot et la pente sur l'UI. Les pentes sont données en demi-tons, puis on calcule un pourcentage.

- *EcMASUT_UI* (écart entre *MAS_UT* et *MAS_UI*) : comme au-dessus, mais l'écart est mesuré entre l'UT et l'UI.

Nous avons fait les mêmes calculs que pour les deux dernières variables, mais sur la valeur de la pente sans saut d'octaves (*MASWOJ*) :

- *EcMASWOJmot_UI* (écart entre *MASWOJ_mot* et *MASWOJ_UI*)

- *EcMASWOJUT_UI* (écart entre *MASWOJ_UT* et *MASWOJ_UI*) : ces deux valeurs sont exprimées en demi-tons.

4.2.4 Intensité

Les paramètres que nous avons pris pour l'intensité sont décrits en 3.3. Ils sont largement inspirés de ceux pris pour la fréquence fondamentale. Nous avons 18 paramètres qui sont tous des variables numériques¹⁰³.

- | | |
|------------------------|-----------------------|
| - <i>MinInt_mot</i> ; | - <i>TmaxInt_UT</i> ; |
| - <i>TMinInt_mot</i> ; | - <i>MeanInt_UT</i> ; |
| - <i>MaxInt_mot</i> ; | - <i>SDInt_UT</i> ; |
| - <i>TmaxInt_mot</i> ; | - <i>MinInt_UI</i> ; |
| - <i>MeanInt_mot</i> ; | - <i>TMinInt_UI</i> ; |
| - <i>SDInt_mot</i> ; | - <i>MaxInt_UI</i> ; |
| - <i>MinInt_UT</i> ; | - <i>TmaxInt_UI</i> ; |
| - <i>TMinInt_UT</i> ; | - <i>MeanInt_UI</i> ; |
| - <i>MaxInt_UT</i> ; | - <i>SDInt_UI</i> . |

Comme pour la fréquence fondamentale, nous avons ajouté quelques paramètres issus de calculs réalisés à partir des 18 paramètres ci-dessus, toujours largement inspirés de ceux calculés pour la fréquence fondamentale :

- *EcMxInt_MeanIUI* (écart entre *MaxInt_mot* et *MeanInt_UI*) : il s'agit de l'écart relatif entre la valeur maximale d'intensité sur le mot et l'intensité moyenne de l'UI, exprimé en pourcentage.

¹⁰³ Cf 3.3.3 pour les détails.

- *EcMxIUT_MeanIUI* (écart entre *MaxInt_UT* et *MeanInt_UI*) : de même, mais on a pris la valeur maximale de l'intensité sur l'UT.
- *EcMinImt_MeanIUI* (écart entre *MinInt_mot* et *MeanInt_UI*) : on compare ici la valeur minimale de l'intensité sur le mot à l'intensité moyenne de l'UI.
- *EcMinIUT_MeanIUI* (écart entre *MinInt_UT* et *MeanInt_UI*) : de même mais on a remplacé le mot par l'UT.
- *EcMxIntmot_UI* (écart entre *MaxInt_mot* et *MaxInt_UI*) : on compare les valeurs maximales de l'intensité sur le mot et sur l'UI.
- *EcMxIntUT_UI* (écart entre *MaxInt_UT* et *MaxInt_UI*) : on compare les valeurs maximales de l'intensité sur l'UT et sur l'UI.
- *EcMinImot_UI* (écart entre *MinInt_mot* et *MinInt_UI*) : on compare les valeurs minimales cette fois, entre le mot et l'UI ici, puis entre l'UT et l'UI pour le paramètre suivant :
- *EcMinIUT_UI* (écart entre *MinInt_UT* et *MinInt_UI*).
- *EcMeanImot_UI* (écart entre *MeanInt_mot* et *MeanInt_UI*) : de la même manière, on calcule l'écart relatif entre les intensités moyennes du mot et de l'UI ici, puis de l'UT et de l'UI, toujours en pourcentage :
- *EcMeanIUT_UI* (écart entre *MeanInt_UT* et *MeanInt_UI*).

4.2.5 Pauses

Il est très difficile d'intégrer des paramètres pour les pauses dans le tableau de données que nous constituons pour l'analyse statistique. En effet, lorsque nous déterminons un paramètre, il vaut pour tous les mots du corpus, quel que soit leur coefficient d'emphasis. Il est du fait impossible d'inclure le paramètre « durée d'une pause » par exemple.

Nous avons contourné la difficulté en insérant de façon automatique¹⁰⁴ les deux paramètres suivants :

- *mot_prec* (mot précédent) : si ce qui précède est un mot, le programme affiche « mot » ; si c'est une pause, il affiche « pause ».

¹⁰⁴ Grâce à une manipulation réalisée sur Excel.

- *mot_suiv* (mot suivant) : on fait la même manipulation, mais à propos de ce qui suit le mot.

Ainsi, grâce à ces deux paramètres, de type catégoriel, le logiciel de statistiques nous indiquera si le fait d'avoir une pause avant ou après un mot est un indice d'emphase, ou bien peut-être associera-t-il pause et allongement dans certains cas, nous verrons cela.

4.2.6 Amplitude totale

Enfin nous nous sommes inspirée de Beckman (1986) pour le paramètre que nous nommons à son instar amplitude totale (« total amplitude ») :

The total amplitude was an approximation of the intensity integral.
(Beckman, 1986 : 151)

L'intensité sonore est une puissance surfacique, c'est-à-dire une valeur ponctuelle $I(t)$ à un instant 't', exprimée soit en W/m^2 , soit en déciBels absolus¹⁰⁵. « L'amplitude totale » intègre cette fonction avec le temps et correspond à l'énergie dispensée pour produire le son.

Cela mesure la surface quadrillée sur la figure 33, surface qui correspond à l'intégrale de l'intensité sur une période de temps déterminée (de a à b), c'est-à-dire le produit de l'intensité instantanée ($I(t)$) par l'intervalle différentiel de temps (dt) :

$$E = \int_a^b I(t)dt$$

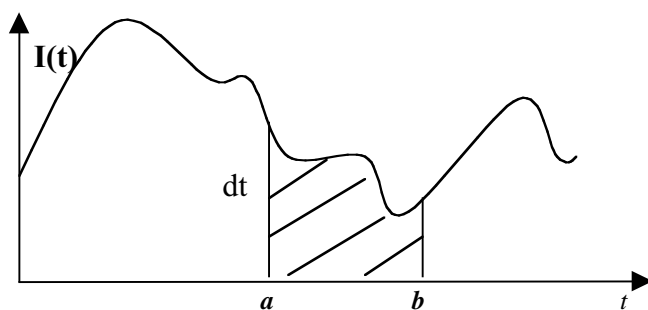


Figure 33 : amplitude totale

¹⁰⁵ Les déciBels absolus sont l'expression logarithmique de l'intensité sonore qui se mesure en $Watt/m^2$. Cf. par exemple : <<http://www.ac-montpellier.fr/scphysiques/SP16/SP161.htm>>

E est l'énergie surfacique, que l'on peut exprimer selon plusieurs unités : Joule/m² ou encore dB.s (¹⁰⁶). Pour notre étude, il est préférable de choisir cette dernière unité, car l'échelle logarithmique est mieux adaptée à l'étendue de la perception de l'oreille humaine.

Ce paramètre est pertinent selon Beckman pour les langues comme l'anglais (« stress accent languages »), où plusieurs paramètres entrent en jeu dans la perception de l'accent mélodique, par opposition au japonais par exemple (« pitch accent language »), où seule la mélodie intervient.

Nous avons donc relevé cette valeur (*TA_mot*) pour chaque mot de notre corpus en effectuant le calcul suivant : *dur_reelle*MeanInt_mot*, ce qui donne une approximation de la surface mentionnée plus haut.

4.2.7 Validation

Comme nous l'avons dit en 4.1.1, il est possible de lancer CRUISE et de valider ensuite les résultats sur un fichier test. Ceci veut dire que nous devons mettre de côté une partie de notre corpus et la réserver pour cette validation. Les auteurs de CRUISE préconisent de faire la validation sur 1/10^e du corpus (cf Kim & Loh, 2001). Pour choisir les mots à mettre de côté, nous avons attribué un groupe de 0 à 9 à chaque mot de notre corpus (les mots sont les uns à la suite des autres dans l'ordre des segments). Ainsi, il ne nous restait plus qu'à extraire un groupe et garder les neuf autres.

Afin d'obtenir des résultats optimaux, nous pouvons lancer CRUISE sur les groupes 1 à 9 et valider sur le groupe 0, puis prendre les groupes 0 à 8 et valider sur le groupe 9, puis prendre les groupes 2 à 0 et valider sur le groupe 1, et ainsi de suite. De cette manière, tout le corpus sera testé. On pourra ensuite faire une synthèse des résultats et choisir les paramètres récurrents.

La variable *groupe* figurera donc dans nos paramètres mais sera bien entendu exclue de l'analyse.

¹⁰⁶ Ce sont des déciBels absolus que multiplient des secondes. En réalité, nous avons multiplié des dB par des millisecondes, ce qui nous donne des dB.ms, que nous diviserons par 1000 pour obtenir des dB.s.

4.2.8 Locuteurs et passages

Nous pensons, comme nous l'avons dit plus haut (chapitre 3), que l'emphase est peut-être rendue de façon différente par chaque locuteur. Il se peut que chacun ait sa stratégie propre ou ses habitudes, et c'est la raison pour laquelle nous avons inclus le paramètre *Locuteur*. Ce paramètre nous servira surtout pour les fichiers de données. Nous utiliserons CRUISE sur les données de chaque locuteur pris séparément, pour voir quels paramètres sont les mieux corrélés à l'emphase chez chacun. Nous comparerons ensuite les résultats.

Nous ferons de même pour ce qui concerne les passages. Nous avons déjà parlé de types de discours différents :

- débat politique (P1), plus préparé peut-être, où l'on cherche à convaincre,
- conversation informelle (P2) sans désaccord,
- émission radio (P3), dont le sujet est très personnel, plus émotionnel que pour les deux autres.

Nous pourrions ainsi voir si l'emphase est rendue différemment selon le type de discours.

4.2.9 Résumé : fichier descriptif

Pour résumer, nous montrons le fichier descriptif que nous obtenons. Comme indiqué plus haut, le « n » signifie numérique, le « c » catégoriel, le « x » exclu et le « d » indique la variable dépendante. Cette dernière peut être différente, d'un essai à l'autre. On peut également décider d'exclure certains paramètres ou d'en inclure d'autres. Le fichier que nous montrons est un exemple, il n'y a pas de version définitive puisqu'on utilisera CRUISE plusieurs fois, sur des fichiers de données différents, avec des variables dépendantes différentes, etc....

La première ligne donne le fichier de données que CRUISE doit analyser. La deuxième indique le symbole pour les valeurs manquantes (le point d'interrogation ici). La troisième ligne indique ce à quoi correspondent les colonnes qui suivent, à savoir le numéro de la colonne du fichier de données, le nom de la variable, et le type de

variable. Le tableau suivant montre un de nos fichiers descriptifs *.dsc* dans la forme exacte requise par CRUISE¹⁰⁷ :

```
DD.dat
?
column, varname, vartype
1      groupe          x          34      TMinF0_UT          x
2      MOT             x          35      MaxF0_UT            x
3      mot_prec        c          36      TMaxF0_UT            x
4      mot_suiv        c          37      Mean_UT             n
5      Locuteur        c          38      SD_UT              n
6      Passage         c          39      MAS_UT             n
7      Segment         x          40      MASWOJ_UT          n
8      UI              x          41      MinF0_UI            x
9      UT              x          42      TMinF0_UI           x
10     %_emph          x          43      MaxF0_UI            x
11     coef_emph1      x          44      TMaxF0_UI           x
12     coef_emph2      x          45      Mean_UI             n
13     coef_emph3      d          46      SD_UI              n
14     %_degr_emph     x          47      MAS_UI             n
15     coef_degr_emph  x          48      MASWOJ_UI          n
16     nb_syll_mot     x          49      EcMxF0mt_MeanUI    n
17     nb_syll_avt_acc x          50      EcMxF0UT_MeanUI    n
18     debut_mot       x          51      EcMxF0mot_UI       n
19     fin_mot         x          52      EcMxF0UT_UI        n
20     dur_reelle_mot  x          53      EcMnmt_MeanF0UI    n
21     dur_predite_mot x          54      EcMnUT_MeanF0UI    n
22     allong_ms_mot   x          55      EcMinF0mot_UI      n
23     allong_%_mot    n          56      EcMinF0UT_UI       n
24     MinF0_mot       x          57      EcMeanF0mot_UI     n
25     TMinF0_mot      x          58      EcMeanF0UT_UI      n
26     MaxF0_mot       x          59      EcMASmot_UI        n
27     TMaxF0_mot      x          60      EcMASUT_UI         n
28     Mean_mot        n          61      EcMASWOJmot_UI     n
29     SD_mot          n          62      EcMASWOJUT_UI      n
30     MAS_mot         n          63      MinInt_mot         x
31     MASWOJ_mot      n          64      TMinInt_mot        x
32     RF_mot          c          65      MaxInt_mot         x
33     MinF0_UT        x          66      TMaxInt_mot        x
```

¹⁰⁷ Le fichier de données correspondant, DD.dat est disponible sur le CD ROM joint, comme tous les autres fichiers de données : *CRUISE : Fichiers .dat*. La liste des abréviations pour les variables utilisées avec CRUISE est donnée en annexe 4a.

67	MeanInt_mot	n	80	SDInt_UI	n
68	SDInt_mot	n	81	EcMxInt_MeanIUI	n
69	MinInt_UT	x	82	EcMxIUT_MeanIUI	n
70	TMinInt_UT	x	83	EcMnInt_MeanIUI	n
71	MaxInt_UT	x	84	EcMnIUT_MeanIUI	n
72	TMaxInt_UT	x	85	EcMxIntmot_UI	n
73	MeanInt_UT	n	86	EcMxIntUT_UI	n
74	SDInt_UT	n	87	EcMinImot_UI	n
75	MinInt_UI	x	88	EcMinIUT_UI	n
76	TMinInt_UI	x	89	EcMeanImot_UI	n
77	MaxInt_UI	x	90	EcMeanIUT_UI	n
78	TMaxInt_UI	x	91	TA_mot	n
79	MeanInt_UI	n			

Tableau 2 : fichier descriptif *emph.dsc*

4.3 Essais préliminaires

Comme nous l'avons déjà expliqué, CRUISE est un logiciel très complexe qui propose des valeurs par défaut mais qui permet également de faire certains choix. Nous nous sommes livrée à quelques essais pour déterminer la meilleure technique et parvenir aux meilleurs résultats possibles pour notre analyse.

4.3.1 SE rule

Comme l'expliquent Kim & Loh (sous presse), cette règle intervient au moment de l'élagage de l'arbre : si l'on choisit 0-SE, on aura plus de chances d'avoir des prédictions pour chaque classe :

[...] if we want every class to be predicted, we have to settle for the more complicated 0-SE tree (Kim & Loh, sous presse : 3)

car en effet avec 1-SE, l'élagage risque d'être trop important :

it can remove so many branches that some classes are not predicted. (*ibid.*)

En résumé, ce paramètre agit sur deux facteurs de façon antinomique :

- le nombre de branches,
- la quantité de classes prédites.

Toute la recherche d'optimisation de ce programme interactif a consisté à choisir judicieusement ce « SE rule », de façon à obtenir des arbres pas trop mais suffisamment « touffus », sur lesquels on trouve le maximum d'occurrences correctement prédites. Car en effet l'arbre le plus gros ne donne pas nécessairement les meilleures prédictions sur le fichier de validation, au contraire. Si les prédictions sont trop précises, elles ne fonctionnent pas lorsque l'on applique les règles trouvées sur un fichier différent.

Cette étude nous a conduite à déterminer une valeur comprise entre 0 et 1,5 pour une gamme possible de 0 à 50. Nous donnons en annexe 4b un exemple d'arbre non élagué puis élagué pour tout le corpus primaire. Dans le fichier de sortie, le nombre de SE choisi est indiqué.

4.3.2 Valeurs par défaut

Il est nécessaire de rentrer dans les options à chaque essai afin de pouvoir préciser que nous possédons un fichier de validation. Nous choisissons pour les autres options les valeurs par défaut (sauf pour le SE). Nous avons fait des essais avec les trois coefficients d'emphase¹⁰⁸ sur les groupes 6 à 4¹⁰⁹ avec le groupe 5¹¹⁰ pour valider et sur les groupes 1 à 9 avec le groupe 0¹¹¹ en validation.

Le coefficient 1 nous donne six classes d'emphase et les résultats¹¹² pour le fichier emph-1-9 avec pour validation emph-0 sont extrêmement mauvais puisque l'on n'obtient pas d'arbre. CRUISE ne parvient pas à prédire les classes. Pour le fichier emph-6-4 validé sur emph-5, on obtient un petit arbre mais les classes 2, 3 et 5 ne sont pas prédites.

Le coefficient 2 donne onze classes d'emphase et CRUISE ne prédit pas mieux : on n'obtient pas d'arbre non plus, pour aucun des deux fichiers.

Le coefficient 3 réduit l'emphase à quatre classes : dans ce cas, CRUISE donne un arbre relativement petit, et le taux de prédiction à partir du fichier emph-1-9 sur le fichier emph-0 est résumé dans la matrice suivante :

Classification Matrix for Test Data					
		Predicted Class			
		0	1	2	3
Actual class	0	162	1	0	2
	1	41	4	0	2
	2	9	2	0	1
	3	8	1	0	4
	Total obs in Test sample	= 237,			
		# correct = 170			

¹⁰⁸ Cf. 4.2.1.

¹⁰⁹ Le fichier est nommé emph-6-4.dat (cf. CD ROM : *CRUISE : Fichiers .dat*).

¹¹⁰ Fichier emph-5.dat

¹¹¹ Fichiers emph-1-9.dat et emph-5.dat.

¹¹² De même, tous les fichiers de sortie (.out) sont disponibles sur le CD ROM : *CRUISE : Fichiers .out*. Ici les fichiers avec le coefficient 1 sont nommés 1_1emph-1-9.out, 1_1emph-6-4.out, et 1_1emph-0.out et 1_1emph-5.out, ceux avec le coefficient 2 commencent par 1_2emph... et pour le coefficient 3 on les appelle 1_3emph... .

Le taux de prédiction correcte est de 71,7% (170 bonnes prédictions sur 237 échantillons). Mais on remarque que la classe 2 n'est jamais prédite. La classe 3 est correctement prédite à 30% dans un cas, mais on voit que 8 occurrences qui sont réellement en classe 3 sont prédites en classe 0, ce qui représente 61% de ce que l'on pourrait appeler une grosse erreur¹¹³. Ce point est important car on peut se trouver avec un nombre de classements corrects assez élevé, du fait que l'on a beaucoup plus d'occurrences dans la classe 0 que dans toutes les autres classes, mais beaucoup de grosses erreurs. Il est plus grave, par exemple, d'avoir une classe 0 d'emphase classée en classe 3, plutôt qu'en classe 1. C'est donc aussi cela qu'il faut observer.

Le coefficient 3 d'emphase donne les meilleurs résultats, et cela n'est pas étonnant. Plus le nombre de classes est élevé, plus le logiciel peine pour trouver des paramètres communs à chaque classe. Nous estimons que le coefficient 3 est pertinent, avec quatre classes d'emphase, ce qui représente déjà une bonne fourchette. Nous privilégierons donc le coefficient 3 pour les analyses suivantes.

4.3.3 Coûts de classification erronée

Nous venons de mentionner le problème des classements erronés. Si l'on a quatre classes, avec dans l'ordre croissant la classe 0 représentant les non-emphatiques et la classe 3 les plus emphatiques, il est plus grave de classer un élément non emphatique (classe 0) en très emphatique (classe 3) plutôt qu'en peu emphatique (classe 1). De la même manière, il est grave de classer un élément de classe 3 en classe 0 plutôt qu'en classe 2. Il existe une option dans CRUISE qui permet d'intégrer cette notion de « gravité ». Elle porte le nom de « misclassification costs » et permet de définir des coûts plus ou moins élevés selon l'erreur de classification. Le logiciel attribue par défaut un coût de 1, quelle que soit l'erreur. On peut modifier les coûts selon l'erreur et rentrer une valeur de 0 à 100.

Nous avons utilisé cette fonction et avons là encore effectué plusieurs essais que nous numérotions. A chaque fois nous utilisons le fichier de validation afin de pouvoir

¹¹³ Les résultats pour le fichier emph-6-4 vont tout à fait dans le même sens.

interpréter plus aisément les résultats¹¹⁴. Nous n'avons effectué ces essais que sur deux fichiers et leurs fichiers test, les mêmes que ci-dessus.

(1) L'essai (1) correspond à ce que nous avons fait en 4.3.2, avec les coûts de classification erronée égaux. Les résultats sont ceux commentés ci-dessus (en 4.3.2) pour le coefficient 3.

(2) Nous avons calculé la différence entre les classes et pris cette valeur. Ainsi, si l'erreur est de classer la classe 0 en 3, cela a un coût de 3. Si c'est une erreur de la classe 2 à la classe 1, le coût est de 1, etc... Pour plus de clarté, cela peut aussi s'exprimer grâce à une matrice semblable à celle des prédictions (CRUISE donne cette matrice dans le fichier de sortie dès que l'on choisit l'option « unequal misclassification costs ») :

Cost selection: Unequal misclassification costs				
Actual class	Predicted Class			
	0	1	2	3
0	0.000	1.000	2.000	3.000
1	1.000	0.000	1.000	2.000
2	2.000	1.000	0.000	1.000
3	3.000	2.000	1.000	0.000

Les résultats à partir du fichier emph-1-9 sur le fichier test emph-0 donnent un arbre très complexe mais seule la classe 0 est prédite dans le fichier de validation. Pour emph-6-4, les classes 1 et 2 sont prédites, mais toujours pas la classe 3.

(3) Nous avons choisi d'augmenter légèrement les écarts et de prendre la différence (comme dans (2)) mais élevée au carré, afin de prendre une échelle non plus linéaire, mais logarithmique cette fois :

¹¹⁴ Tous les résultats sont consultables sur le CD ROM : *CRUISE : Fichiers .out*. Ici les fichiers sont numérotés comme les essais : l'essai 3 sur les groupes 1 à 9 par exemple sera nommé 3emph-1-9.out (sauf pour l'essai (1) qui reste nommé comme en 4.3.2 : 1_3emph...)

Cost selection: Unequal misclassification costs				
Actual class	Predicted Class			
	0	1	2	3
0	0.000	1.000	4.000	9.000
1	1.000	0.000	1.000	4.000
2	4.000	1.000	0.000	1.000
3	9.000	4.000	1.000	0.000

Les fichiers de sortie montrent que CRUISE a bien du mal à prédire la classe 3 :

Classification Matrix for Test Data emph-0				
Act class	Predicted Class			
	0	1	2	3
0	127	34	4	0
1	15	28	4	0
2	2	6	4	0
3	3	5	5	0
Test sample = 237, # correct = 159				

Classification Matrix for Test Data emph-5				
Act class	Predicted Class			
	0	1	2	3
0	114	39	3	0
1	15	21	5	0
2	5	8	6	0
3	3	11	8	0
Test sample = 238, # correct = 160				

(4) Plutôt que de prendre la différence élevée au carré, nous avons pris la racine carrée de la différence, ce qui fait des écarts plus petits : on prend 1 pour une différence de 1, $\sqrt{2}$ (1,414) pour une différence de 2 et $\sqrt{3}$ (1,732) pour une différence de 3.

Les résultats sont encourageants pour le fichier emph-1-9 validé sur emph-0, car toutes les classes sont prédites. Mais dans notre deuxième cas, la classe 2 ne sort pas.

Classification Matrix for Test Data emph-0				
Act class	Predicted Class			
	0	1	2	3
0	150	13	1	1
1	31	11	2	3
2	3	6	2	1
3	5	3	2	3
Test sample = 237, # correct = 166				

Classification Matrix for Test Data emph-5				
Act class	Predicted Class			
	0	1	2	3
0	144	11	0	1
1	27	12	0	2
2	8	5	0	6
3	6	7	0	9
Test sample = 238, # correct = 165				

On remarque en outre que les erreurs ne sont pas trop graves pour emph-0, mais on a encore 5 occurrences de classe 3 classées en classe 0, sur 13.

Nous avons continué nos recherches.

(5) Nous avons jusque-là utilisé des coûts de classification erronée symétriques. Nous avons voulu voir ce que cela donnerait si l'on accordait moins d'importance au fait de classer des « 3 » en « 0 » plutôt que des « 0 » en « 3 ». L'argument est que comme CRUISE a du mal prédire la classe 3, celle-ci est rare et nous aurons donc nettement moins d'erreurs de « 3 » à « 0 » que de « 0 » à « 3 ». Nous avons élaboré une matrice dissymétrique basée sur les puissances de 4 pour les erreurs les plus graves, et les racines 4^{ème} pour les erreurs les moins graves avec interpolation linéaire :

Cost selection: Unequal misclassification costs					
		Predicted Class			
		0	1	2	3
Actual class	-----				
0		0.000	1.000	1.190	1.320
1		40.000	0.000	1.000	1.190
2		60.000	40.000	0.000	1.000
3		81.000	60.000	40.000	0.000

Nous avons donc augmenté encore les écarts dans un sens et les avons réduits dans l'autre.

Les résultats sont complètement déséquilibrés, CRUISE ne prédit cette fois que la classe 3, ce qui ne donne que très peu de classements corrects (13 sur 237 et 22 sur 238) dans les fichiers de validation :

Classification Matrix for Test Data emph-0					
		Predicted Class			
		0	1	2	3
Act class	-----				
0		0	0	0	165
1		0	0	0	47
2		0	0	0	12
3		0	0	0	13
Test sample = 237, # correct = 13					

Classification Matrix for Test Data emph-5					
		Predicted Class			
		0	1	2	3
Act class	-----				
0		0	0	0	156
1		0	0	0	41
2		0	0	0	19
3		0	0	0	22
Test sample = 238, # correct = 22					

(6) Nous avons essayé d'exacerber la classe 3, qui est finalement la plus importante pour notre étude, et de diminuer le coût d'une erreur d'une classe. Là encore, la matrice de coûts de classification erronée est créée de façon asymétrique :

Cost selection: Unequal misclassification costs				
Actual class	Predicted Class			
	0	1	2	3
0	0.000	1.000	1.190	81.000
1	0.000	0.000	0.000	60.000
2	1.190	0.000	0.000	40.000
3	1.320	1.190	40.000	0.000

Cette fois CRUISE ne réussit à prédire ni la classe 3 ni la classe 2 :

Classification Matrix for Test Data emph-0				
Act class	Predicted Class			
	0	1	2	3
0	163	2	0	0
1	44	3	0	0
2	10	2	0	0
3	9	4	0	0
Test sample = 237, # correct = 166				

Classification Matrix for Test Data emph-5				
Act class	Predicted Class			
	0	1	2	3
0	153	3	0	0
1	38	3	0	0
2	12	7	0	0
3	14	8	0	0
Test sample = 238, # correct = 156				

Le nombre de classements corrects est assez élevé, mais cela est dû au fait qu'il prédit bien la classe 0, alors que nous aurions préféré qu'il prédise bien la classe 3.

Les essais (5) et (6) nous confortent dans l'idée qu'il est souhaitable et plus logique de garder des coûts symétriques. Il est aussi « grave » de mal classer dans un sens que dans l'autre.

(7) Nous en revenons donc à la notion de puissance, comme dans les essais (2) (puissance 1) et (3) (puissance 2) et nous avons essayé puissance 4, donnant des coûts de 1, 16 ou 81. A nouveau, la classe 3 n'est pas prédite.

(8) Nous avons alors essayé puissance 5, avec des coûts de 0.41, 13.16 et 100. Les résultats ne sont toujours pas bons. Peut-être allions-nous dans le mauvais sens ?

(9) Nous avons alors pris puissance 3 avec des coûts de 1, 8 ou 27. Les résultats sont à nouveau à peu près similaires. La classe 3 n'est pas prédite.

(10) Les résultats des essais (3) et (4), qui donnent jusqu'ici les meilleures prédictions en ce qui concerne la classe 3, sont révélateurs car si la classe 3 est prédite, c'est la classe 2 qui est absente. Ceci semblerait vouloir dire que les classes 2 et 3 ne sont pas vraiment différentes. On penche alors vers une notion plus dichotomique de l'emphase. Un mot serait emphatique ou non-emphatique au niveau acoustique. Nous avons pensé qu'il fallait alors creuser l'écart entre la classe 0 et les autres classes. On aurait alors un mélange entre la dichotomie emphatique / non emphatique et des degrés à l'intérieur de l'emphatique (cf. Ladd *et al.*, 1994). Nous avons créé la matrice de coûts de classification erronée suivante :

Cost selection: Unequal misclassification costs				
Actual class	Predicted Class			
	0	1	2	3
0	0.000	25.000	36.000	49.000
1	25.000	0.000	25.000	36.000
2	36.000	25.000	0.000	25.000
3	49.000	36.000	25.000	0.000

A nouveau les classes 2 et 3 ne sont pas prédites simultanément :

Classification Matrix for Test Data emph-0				
Act class	Predicted Class			
	0	1	2	3
0	150	13	2	0
1	32	12	3	0
2	3	7	2	0
3	5	4	4	0
Total obs in Test sample =				
237, # correct = 164				

Classification Matrix for Test Data emph-5				
Act class	Predicted Class			
	0	1	2	3
0	137	18	0	1
1	23	17	0	1
2	6	9	0	4
3	4	12	0	6
Total obs in Test sample =				
238, # correct = 160				

(11) Nous avons encore fait un essai sur le même principe que (10) en changeant légèrement les écarts :

Cost selection: Unequal misclassification costs				
Actual class	Predicted Class			
	0	1	2	3
0	0.000	25.000	49.000	81.000
1	25.000	0.000	4.000	16.000
2	49.000	4.000	0.000	4.000
3	81.000	16.000	4.000	0.000

La classe 3 n'est pas prédite.

Nous pouvons tirer de tous ces essais une conclusion partielle : ce sont les essais (3) et (4) qui nous donnent les meilleurs résultats en ce qui concerne les prédictions pour la classe 3. Mais tout cela n'est pas vraiment convaincant.

La classe 3 étant rarement prédite, nous avons pensé qu'il y avait peut-être trop peu d'écart entre les classes 2 et 3 et avons donc décidé de supprimer une classe. Nous avons donc fait de nouveaux essais en nous basant sur les essais (4) et (10), afin de voir la différence : dans un cas, les degrés d'emphase sont privilégiés, dans l'autre, c'est plutôt le concept de la dichotomie.

4.3.4 Nouveaux coefficients d'emphase

Nous avons déterminé deux nouveaux coefficients d'emphase basés sur un découpage linéaire :

- *coef_emph4* ne comporte que deux classes, basées comme pour les autres coefficients d'emphase sur le pourcentage d'emphase des sondages de perception. La classe 0 regroupe les mots emphatiques entre 0 et 49%, la classe 1 entre 50 et 100%.
- *coef_emph5* comprend trois classes, les mots emphatiques de 0 à 32% pour la classe 0, de 33 à 65% pour la classe 1 et de 66 à 100% pour la classe 2.

Nous avons à nouveau effectué des essais sur les échantillons *emph-1-9* avec son fichier test *emph-0*, et *emph-6-4* et son fichier test *emph-5*.

(a) Le premier essai concerne le coefficient n°4, c'est-à-dire deux classes uniquement, selon le principe de la dichotomie mentionnée plus haut. Comme il n'y a que deux classes, les coûts de classification erronée n'ont plus aucun intérêt, nous avons laissé la valeur par défaut. Les résultats ne nous ont pas paru probants¹¹⁵ :

Classification Matrix for Test Data emph-0				Classification Matrix for Test Data emph-5			
		Predicted Class				Predicted Class	
		0	1			0	1
Act class	0	205	7	Act class	0	194	3
	1	16	9		1	33	8
Test sample = 237, # correct = 214				Test sample = 238, # correct = 202			

Le nombre de mots de classe 1 mis dans la classe 0 nous semble trop important : il est de 64% (16 mots sur 25) et 80% (33 sur 41).

Nous laisserons donc de côté le coefficient 4, qui nous paraît du reste être une trop grande simplification.

Les essais avec le coefficient 5 (trois classes d'emphase) sont plus révélateurs selon nous. Les coûts de classification erronée prennent à nouveau de l'importance.

¹¹⁵ Les détails des résultats sont sur le CD ROM dans *CRUISE : Fichiers .out : emph-1-9.out, emph-6-4.out, emph-0.out et emph-5.out.*

(b) Nous avons pris des coûts de 1 ou 8 selon l'écart entre les classes, sur le même principe que les premiers essais ci-dessus. CRUISE n'arrive pas à prédire la classe 2¹¹⁶. On se retrouve dans la même situation que pour le coefficient 3.

(c) Nous avons pris la puissance 2, comme en (3), les résultats¹¹⁷ sont similaires à ceux de (b).

(d) Ici c'est le principe de (10) que nous avons appliqué, c'est-à-dire que nous avons rentré un coût plus important pour les erreurs de 0 à 1 ou de 0 à 2 et inversement, que pour les erreurs de 1 à 2 ou de 2 à 1 :

Cost selection: Unequal misclassification costs				
		Predicted Class		
		0	1	2
Actual class -----				
	0	0.000	25.000	36.000
	1	25.000	0.000	10.000
	2	36.000	10.000	0.000

Une occurrence de classe 2 est bien prédite dans le fichier de validation, ce qui est encourageant.

(e) Nous avons continué en augmentant les écarts :

Cost selection: Unequal misclassification costs				
		Predicted Class		
		0	1	2
Actual class -----				
	0	0.000	75.000	100.000
	1	75.000	0.000	25.000
	2	100.000	25.000	0.000

Les résultats¹¹⁸ sont pratiquement les mêmes qu'en (d), la classe 2 est cependant moins bien prédite en (e).

¹¹⁶ Cf CD ROM : CRUISE : Fichiers .out : *emphb-1-9.out*, *emphb-6-4.out*, *emphb-0.out* et *emphb-5.out*.

¹¹⁷ Cf CD ROM : CRUISE : Fichiers .out : *emphc-1-9.out*, *emphc-6-4.out*, *emphc-0.out* et *emphc-5.out*

¹¹⁸ Cf CD ROM : CRUISE : Fichiers .out : *emphd-1-9.out*, *emphd-6-4.out*, *emphd-0.out* et *emphd-5.out*, et *emphe-1-9.out*, *emphe-6-4.out*, *emphe-0.out* et *emphe-5.out*

On obtient soit un meilleur nombre de bonnes classifications (d), mais de plus graves erreurs, soit moins de bonnes réponses mais de moins graves erreurs (e). Nous notons tout de même que (e) donne de moins bonnes prédictions pour la classe 2, mais elles sont meilleures pour la classe 1. La différence n'est toutefois pas flagrante et nous avons fait ces deux essais sur d'autres échantillons et en particulier sur des données par locuteur. Les résultats sont meilleurs, la classe 2 est prédite. La différence entre les deux options est faible, c'est souvent une prédiction qui les départage, et pas toujours dans le même sens :

- pour emph-7-5 validé sur emph-6, (d) semble meilleur ;
- pour emph-2-0, (e) est meilleur pour deux occurrences ;
- pour emph-6-4, (e) est meilleur aussi, à une occurrence près.

Lors des essais sur les données par locuteur, nous ne validons pas les données car nous n'en possédons pas suffisamment pour pouvoir en enlever 10%.

- chez YB, (d) est meilleur pour une occurrence,
- chez PM, (e) est meilleur en ce qui concerne la classe 2 et légèrement moins bon pour la classe 1.

Il est clair après tous ces essais que nous ne parvenons pas à trouver des coûts de classification erronée idéaux. Nous supputons que cela est dû à nos données et qu'il n'existe pas d'idéal. Nous avons aussi du mal à trouver un coefficient d'emphase idéal. En effet, nous avons essayé le coefficient 5 sur bon nombre de fichiers et chez MM, par exemple, CRUISE ne sort pas d'arbre. Si l'on prend le coefficient 4 (2 classes d'emphase), on arrive à un arbre.

4.3.5 Résumé et conclusion : discussion sur l'aspect catégoriel de l'emphase

Nous avons déterminé et expliqué toutes les variables que nous rentrons dans le fichier de données.

Nous avons effectué de nombreux essais afin de choisir une formule que nous allons appliquer à tous nos fichiers de données. Les essais étaient nécessaires puisque nous utilisons un outil tout récent que nous expérimentons.

Il est ardu de faire un choix entre les différents essais relatés ci-dessus. Nous pourrions faire le choix de lancer CRUISE avec différentes options et de comparer les paramètres qu'il donne comme essentiels dans la perception de l'emphase.

Nous n'avons pas encore parlé des paramètres et n'avons regardé pour prendre une décision que les résultats des prédictions. Mais tout au long des essais que nous avons effectués, les mêmes paramètres reviennent, de façon légèrement différente et corrélés entre eux différemment aussi, mais on peut dégager une constante. Ceci est encourageant et prouve que quelle que soit l'option choisie, on n'arrivera pas à des conclusions complètement opposées.

Finalement, la discussion entreprise autour des coûts de classification erronée nous ramène à la discussion sur l'aspect catégoriel de l'emphase (cf. Ladd *et al.*, 1994¹¹⁹). Est-ce un concept dichotomique ou existe-t-il plusieurs degrés d'emphase ?

La réponse à cette question semble être : oui, l'emphase est un phénomène dichotomique. On peut dire qu'un mot est emphatique, ou non-emphatique. Le fait que CRUISE ait du mal à prédire la classe la plus emphatique accrédite cette hypothèse.

Mais nous pensons aussi que : oui, il existe plusieurs degrés d'emphase, dans la catégorie emphatique. Les résultats des essais avec seulement trois classes et des coûts de classification comme ceux décrits dans les essais (10) ou (e) corroborent cette idée.

Pour toutes ces raisons et grâce à tous ces essais, nous faisons le choix du coefficient 5 d'emphase avec seulement trois degrés repartis ainsi :

- Classe 0 : non emphatique ;
- Classes 1 et 2 emphatiques, avec la classe 1 moins emphatique que la classe 2.

Pour ce faire nous déterminons les coûts de classification erronée suivants (matrice symétrique) :

Cost selection: Unequal misclassification costs				
		Predicted Class		
		0	1	2
Actual class	0	0.000	75.000	100.000
	1	75.000	0.000	25.000
	2	100.000	25.000	0.000

¹¹⁹ Travaux décrits en 1.2.2.3. et cf. également 2.2.1 pour une discussion sur l'aspect catégoriel ou continu de l'emphase.

Nous avons choisi 100 comme chiffre le plus élevé car c'est le coût maximal autorisé.
Cette matrice est exactement l'équivalente de la matrice suivante :

0	0.000	3.000	4.000
1	3.000	0.000	1.000
2	4.000	1.000	0.000

Nous pouvons désormais sortir des arbres pour tous les fichiers de données que nous possédons.

4.4 Analyses

Nous analyserons dans un premier temps le corpus primaire en entier, en réservant 10% des données pour la validation. Cela nous donne 10 fichiers différents.

Nous utiliserons ensuite CRUISE pour voir si des constantes propres à chaque type de discours se dégagent.

Enfin nous lancerons une analyse statistique sur les données de chaque locuteur afin de voir si des stratégies personnelles apparaissent.

4.4.1 Remarques préliminaires

Nous souhaitons expliquer ici de quelle manière nous allons procéder pour l'interprétation des arbres. Ces derniers sont en effet assez complexes à première vue, mais on peut généralement en faire une bonne synthèse si on les regarde de plus près.

Prenons l'exemple du fichier *emph-2-0*. On obtient des prédictions correctes en élaguant légèrement l'arbre (1 SE), car si on l'élague trop, il est très complexe et les prédictions sur le fichier de validation ne sont pas meilleures¹²⁰.

La figure 34 ci-dessous représente l'arbre obtenu. Le premier nœud reprend les données initiales du fichier : 1713|214|210. Il s'agit du nombre d'occurrences dans chaque classe : 1713 occurrences sont dans la classe 0, 214 dans la classe 1 et 210 dans la classe 2. Ensuite, chaque arborescence est relative à un paramètre qui est inscrit à la gauche de la ligne. Dans notre exemple, la première arborescence est relative à *TA_mot* et comprend deux branches. Sous l'intitulé du paramètre, on trouve l'intervalle auquel correspond le nœud : ici par exemple, les deux nœuds divisent les occurrences en deux groupes, celles dont l'amplitude totale est inférieure ou égale à 55854,8 dB.ms et celles dont l'amplitude totale est supérieure à 55854,8 dB.ms. Si l'on observe la quatrième ligne de nœuds (sans compter le premier noeud), on voit qu'il y a une séparation en trois :

- $TA_{mot} \leq 15364,7$;
- $TA_{mot} \leq 21341,9$;
- $TA_{mot} > 21341,9$.

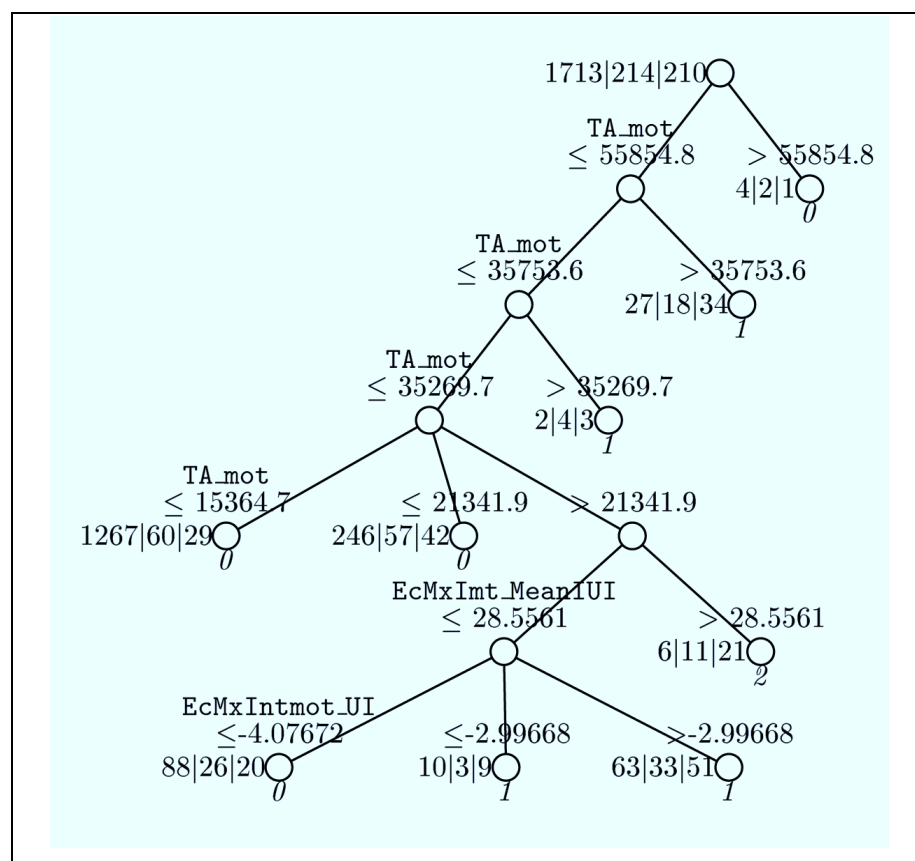


Figure34 :emph-2-0, arbre correspondant au fichier emph-2-0.out

Le nœud central regroupe en fait les occurrences dont l'amplitude totale est comprise entre 15364,7 et 21341,9 dB.ms. Pour les deux autres nœuds, les valeurs inférieure (pour le nœud de gauche) et supérieure (pour celui de droite) sont 0 et l'infini respectivement.

Lorsqu'un nœud est terminal, le nombre d'occurrences dans chaque classe est donné (les chiffres sont séparés par des barres), et le nombre sous le nœud indique la classe prédite. Dans notre exemple, trois classes existent (0, 1 et 2), puisque le coefficient d'emphase est le n°5.

Toujours sur la quatrième ligne de nœuds, mentionnée juste au-dessus, il est possible en fait de regrouper les deux nœuds de gauche (qui sont tous deux terminaux) puisque le

¹²⁰ Cf. 4.3.1.

même paramètre entre en jeu et que les deux nœuds prédisent la même classe (0 ici). On peut également regrouper les deux nœuds qui prédisent la classe 1 à droite (lignes 2 et 3 sans compter le premier nœud). Si l'on s'arrête à notre quatrième ligne, on peut dire que :

- a) Si $TA \leq 21341,9$ dB.ms : classe 0. On a regroupé les deux nœuds. On se trouve sur la quatrième ligne.
- b) Si TA est comprise entre 21341,9 et 35269,7 dB.ms : classes 0, 1 ou 2 (on regarde pour voir cela les nœuds terminaux qui se trouvent en-dessous). Pour départager les classes, il faut descendre aux nœuds suivants, et de nouveaux paramètres entrent en jeu. Nous y reviendrons.
- c) Si TA est comprise entre 35269,7 et 55854,8 dB.ms : classe 1. On a encore regroupé deux nœuds.
- d) Si $TA > 55854,8$ dB.ms : classe 0.

Reprenons le cas c) : la classe 1 est prédite (2^{ème} et 3^{ème} lignes, nœuds terminaux de droite), mais si l'on regarde le nombre d'occurrences dans chaque classe, on remarque qu'il y a pour le nœud terminal de la 2^{ème} ligne 27 occurrences qui appartiennent à la classe 0, 18 à la classe 1 et 34 à la classe 2. Alors pourquoi est-ce la classe 1 qui est prédite ? Ceci est dû aux coûts de classification erronée. Le logiciel fait un calcul en fonction de cela :

- s'il prédit la classe 0, les coûts seront de $18(\text{nombre d'occurrences en classe 1}) \times 75 (\text{coût d'une erreur de classe 1 classée en 0}) = 1350$. A ceci il faut rajouter $34(\text{nombre d'occurrences en classe 2}) \times 100 (\text{coût d'une erreur de classe 2 classée en 0}) = 3400$. Ce qui fait un coût de 4750.
- s'il prédit la classe 2, les mêmes calculs donnent un coût de $(27 \times 100) + (18 \times 25) = 3150$.
- s'il prédit la classe 1, le coût est de $(27 \times 75) + (34 \times 25) = 2875$.

CRUISE choisit donc la classe pour laquelle le coût d'erreur est le moins élevé, même si c'est celle qui comporte le moins d'occurrences, comme c'est le cas pour la classe 1 ici.

Cette remarque nous semble essentielle, car on pourra nuancer les résultats donnés par CRUISE grâce à ces chiffres révélateurs. Un autre exemple illustre ce point : regardons la dernière ligne de nœuds. Le nœud de droite est terminal et prédit la classe 1 en

fonction d'une nouvelle variable : *EcMxIntmot_UI*. A nouveau, on remarque qu'il y a davantage d'occurrences dans la classe 2 que dans la classe 1, et c'est pourtant la classe 1 qui est prédite. Cela est dû au fait qu'il y en a également un nombre assez important dans la classe 0.

Reprenons le cas b) ci-dessus : si *TA* est comprise entre 21341,9 et 35269,7 dB.ms, les classes prédites sont 0, 1 ou 2. Pour les différencier, on doit rajouter le paramètre *EcMxIntmot_MeanUI* :

- si $EcMxIntmot_MeanUI \leq 28,55\%$: classes 0 ou 1. Il faut encore rajouter un paramètre qui est *EcMxIntmot_UI*. Si celui-ci est inférieur ou égal à -4,07%, la classe 0 est donnée. S'il est supérieur ou égal à -2,99%, c'est la classe 1. On a regroupé les deux derniers nœuds de droite qui prédisent tous deux la classe 1.

- si $EcMxIntmot_MeanUI > 28,55$: classe 2.

Notons une dernière chose : sur la première ligne de nœuds, on voit que si $TA > 55854,8$ dB.ms, on obtient la classe 0. Cette prédiction repose sur 7 occurrences, 4 en classe 0, 2 en classe 1 et 1 en classe 2. Si l'on regroupe ce nœud avec les deux nœuds de droite qui se trouvent dessous (que l'on a déjà regroupés), on additionne le nombre d'occurrences dans chaque classe et on arrive aux chiffres suivants : si $TA > 35269,7$ dB.ms, on a 33 occurrences en classe 0, 24 en classe 1 et 38 en classe 2. D'après les coûts de classification erronée choisis, cela donne 5600 points d'erreur si l'on classe en 0, 3425 points si l'on classe en 1 et 3900 points si l'on classe en 2. La prédiction sera donc la classe 1.

Il est important de regarder ces chiffres, car nous avons déterminé les coûts de classification en fonction d'une notion dichotomique de l'emphase avec deux degrés dans la catégorie emphatique. Dans cet exemple, il semble difficile de distinguer les classes 1 et 2, et nous pouvons donc dire que si $TA > 35269,7$, l'occurrence va être en classe 1 ou 2, donc emphatique.

On peut résumer et synthétiser cet arbre de la façon suivante :

si $TA \leq 21341,9$ dB.ms : classe 0.

si TA entre 21341,9 et 35269,7 dB.ms, **et** $EcMxIntmot_MeanUI \leq 28,55\%$ **et** $EcMxIntmot_UI \leq -2,99\%$: classe 0.

si TA entre 21341,9 et 35269,7 dB.ms, **et** $EcMxIntmot_MeanUI \leq 28,55\%$ **et**

$EcMxIntmot_UI > -2,99\%$: classe 1 (ou 2).

si TA entre 21341,9 et 35269,7 dB.ms, **et** $EcMxIntmot_MeanUI > 28,55\%$: classe 2.

si $TA > 35269,7$ dB.ms : classe 1(ou 2).

C'est ainsi que nous procéderons pour chaque arbre, et nous ne détaillerons plus aussi précisément nos analyses. Nous rentrerons par contre dans le détail des paramètres et de ce à quoi ils correspondent, chose que nous n'avons volontairement pas entreprise ici. Nous souhaitons uniquement, dans cette sous-section, éclaircir la façon dont nous analysons les arbres.

4.4.2 Corpus primaire

Comme indiqué ci-dessus, nous avons demandé à CRUISE des arbres pour le corpus primaire en enlevant 10% des données afin d'obtenir des prédictions de classification sur des fichiers de validation. Nous avons donc fait ce travail sur dix fichiers. Les arbres sont de taille très variable. Nous avons choisi différentes options d'élagage, et sélectionné pour chaque fichier celle qui donnait les meilleurs résultats sur les fichiers de validation. Tous les arbres ne résultent donc pas du même élagage. Le nombre de SE varie de 0 à 1,5 (cf. 4.3.1).

Les prédictions correctes obtenues sur les fichiers de validation sont intéressantes à analyser. Elles varient considérablement d'un fichier à l'autre, mais nous pouvons dégager une constante :

- la classe 0 est correctement prédite à 94,6% en moyenne sur tous les fichiers. Ceci n'est pas très étonnant car c'est dans cette classe que l'on trouve le plus grand nombre d'occurrences ;
- la classe 1 est correctement prédite à 18,5% de moyenne sur les 10 fichiers ;
- la classe 2 est correctement prédite à 11% de moyenne ;

Même si ces résultats peuvent sembler médiocres en ce qui concerne la classe 2 (la plus emphatique) en particulier, cela veut dire que l'on arrive tout de même à des prédictions correctes autour de 30% pour la classe emphatique, et 95 % pour la classe non emphatique, ce qui nous semble pertinent.

Quant aux taux d'erreurs entre la classe 0 et la classe 2, ils sont comme suit :

- la moyenne pour les erreurs d'occurrences de classe 2 ou 1 classées en classe 0 est de 56,5 % ;

- la moyenne pour les occurrences de classe 0 classées en 2 est de 1%.

Le premier taux est mauvais, mais le second excellent.

Les arbres pour chacun des dix fichiers étant différents¹²¹, la solution pour en faire une synthèse était de lancer CRUISE sur le corpus tout entier, sans fichier de validation cette fois, d'où une difficulté pour choisir le bon élagage. Sans aucun élagage, l'arbre est très gros, et nous avons vu qu'il n'est pas forcément meilleur, car CRUISE se base sur très peu d'occurrences pour prédire une classe. Au vu des résultats sur le fichier lui-même et grâce à l'expérience acquise au cours des essais effectués sur les dix autres fichiers, nous avons opté pour l'élagage avec 0,8 SE¹²².

La figure 35 ci-dessous présente l'arbre pour tout le corpus. C'est celui que nous proposons d'interpréter, auquel nous ajouterons quelques observations découlant de l'analyse des dix fichiers mentionnés précédemment.

D'après les explications données ci-dessus, voici une synthèse de cet arbre :

- A)** si $TA > 35355,1$ dB.ms : classe 1 (ou 2) ;
si $TA \leq 15612,2$ dB.ms : classe 0 ;
- B)** si TA entre 15612,2 et 21470,7 dB.ms, et $EcMeanF0mot_UI \leq 38,13\%$: classe 0 ;
si TA entre 15612,2 et 21470,7 dB.ms, et $EcMeanF0mot_UI > 38,13\%$: classe 2 ;
- C)** si TA entre 21470,7 et 35355,1 dB.ms, et $EcMxIntmot_MeanUI > 30,46\%$: classe 2 ;
si TA entre 21470,7 et 35355,1 dB.ms, et $EcMxIntmot_MeanUI \leq 30,46\%$ et $MASWOJ_mot \leq 51,49$ dt/s et $EcMxIntmot_UI \leq -4,04\%$: classe 0 ;

¹²¹ Les fichiers de sortie dans lesquels figurent tous les détails des options choisies sont disponibles sur le CD ROM : *CRUISE : Fichiers .out*. Les arbres pour les fichiers finalement sélectionnés, ainsi que le nombre de SE et la matrice de confusion pour le fichier test, sont présentés en annexe 4b.

¹²² Sur le CD ROM joint, tous les fichiers sont disponibles : *emph.out* est l'essai avec 0SE, *emphbis.out* correspond à 1SE, *emphter.out* à 0,5SE et *emphqat.out* à 0,8SE. Cf. CD ROM : *CRUISE : Fichiers .out*.

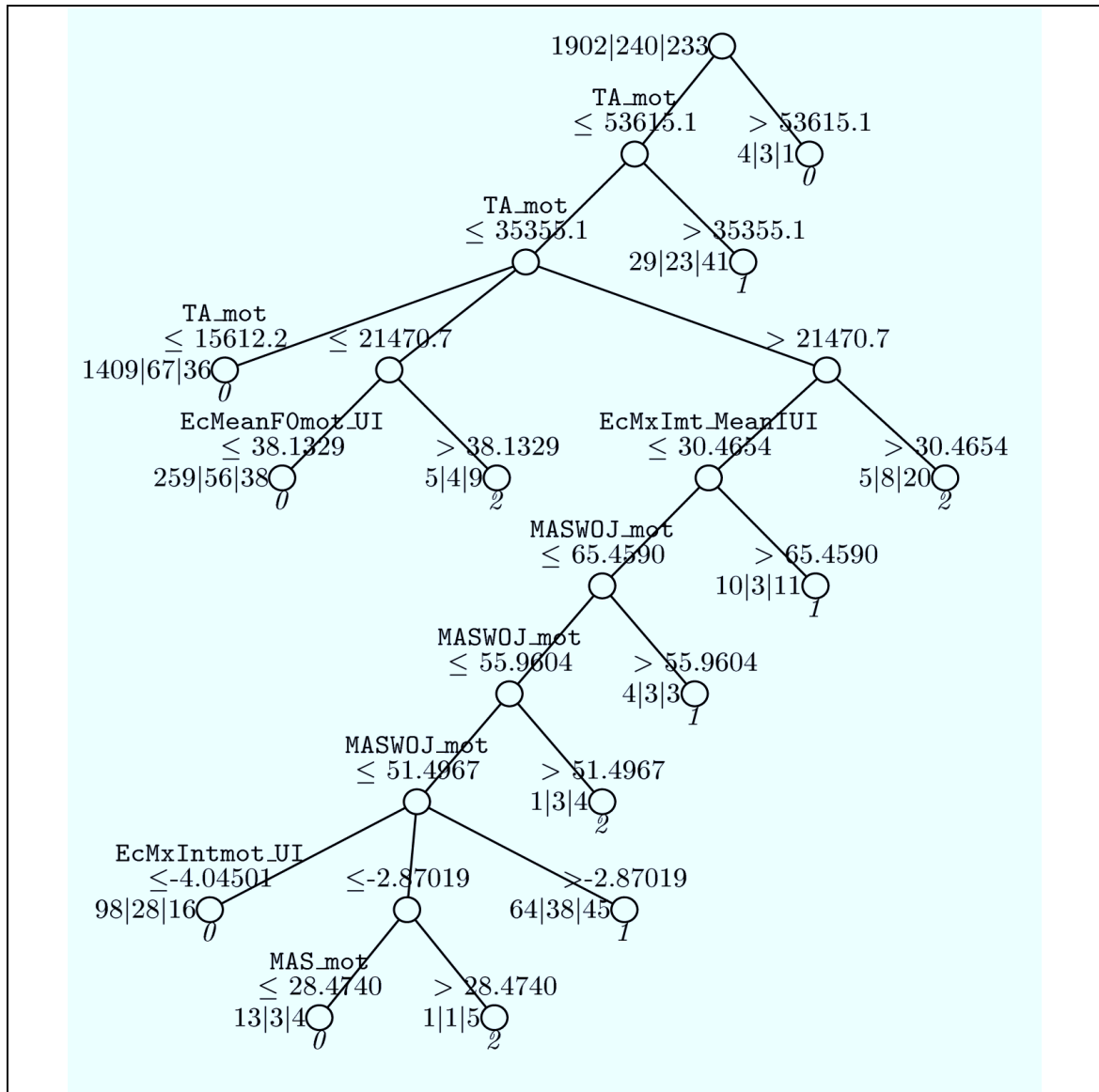


Figure 35 : arbre de régression pour tout le corpus primaire

si TA entre 21470,7 et 35355,1 dB.ms, et $EcMxIntmot_MeanUI \leq 30,46\%$ et $MASWOJ_mot \leq 51,49$ dt/s et $EcMxIntmot_UI$ entre -4,04% et -2,87% et $MAS_mot \leq 28,47$ dt/s : classe 0 ;

si TA entre 21470,7 et 35355,1 dB.ms, et $EcMxIntmot_MeanUI \leq 30,46\%$ et $MASWOJ_mot \leq 51,49$ dt/s et $EcMxIntmot_UI$ entre -4,04% et -2,87% et $MAS_mot > 28,47$ dt/s : classe 2 ;

si TA entre 21470,7 et 35355,1 dB.ms, et $EcMxIntmot_MeanUI \leq 30,46\%$ et $MASWOJ_mot \leq 51,49$ dt/s et $EcMxIntmot_UI > -2,87\%$: classe 1 (ou 2).

Analysons cet arbre en détaillant maintenant les paramètres.

A) Le premier critère, TA , est exprimé en dB.ms. On peut le convertir facilement en dB.s (par exemple $35355,1 \text{ dB.ms} = 35,35 \text{ dB.s}$). Il exprime une énergie (cf. 4.2.6), et nous avons vu que la durée et l'intensité sont prises en compte dans ce paramètre. Si l'amplitude totale est élevée (au-dessus de $35,35 \text{ dB.s}$), il y a emphase. Cela veut dire que pour mettre un mot en relief de façon emphatique, on peut :

- soit prononcer une occurrence courte avec une forte intensité,
- soit rallonger l'occurrence et ne pas trop l'intensifier,
- soit associer les deux, la longueur et l'intensité.

L'inverse est vrai puisque en dessous de $15,6 \text{ dB.s}$, l'occurrence ne sera pas du tout emphatique. Ainsi si l'on prononce une occurrence sans l'intensifier ou en la raccourcissant, ou les deux, elle ne sera pas emphatique.

Prenons quelques exemples pour illustrer cela.

Dans P2.3S01, CB emploie un mot assez court : « home ». Mais elle rallonge ce mot de 132% par rapport à la durée moyenne des phonèmes et le prononce avec une moyenne de $60,6 \text{ dB}$, ce qui est assez fort. Cela donne une amplitude totale de $36,33 \text{ dB.s}$, indice d'emphase ici.

Dans le cas de mots longs, comme « liberates » (P3.1S08), donné emphatique par 100% des auditeurs, l'amplitude totale est de 43 dB.s . En effet, sa durée est de $754,2 \text{ ms}$ et son intensité moyenne de 57 dB . Ce dernier chiffre est assez élevé, mais pas démesurément élevé. C'est la durée qui rend l'amplitude totale si haute.

Mais ceci pose un problème : si un mot est excessivement long (s'il contient trois ou quatre syllabes par exemple), il a de grandes chances d'avoir une amplitude totale supérieure à $35,35 \text{ dB.s}$ sans pour autant être emphatique. C'est la raison pour laquelle on trouve bon nombre d'occurrences de classe 0 dont l'amplitude totale est supérieure à $35,35 \text{ dB.s}$. Il faut donc ici nuancer notre propos et préciser que cela fonctionne en règle générale, à l'exception de mots très longs. De même, si un mot très court est emphatique, il risque de se trouver quand même avec une amplitude totale assez faible. C'est pourquoi on a 36 occurrences dont l'amplitude totale est inférieure à $15,6 \text{ dB.s}$ qui se trouvent en classe 2. C'est le cas dans P2.2S05 par exemple : VB met le mot « pop » en relief. Il est donné emphatique par 75% des auditeurs, et son amplitude totale n'est

que de 15,45 dB.s. Sa durée n'est que de 292 ms, c'est un mot court, et son intensité moyenne est de 54 dB du fait qu'il comprend deux consonnes sourdes. Dans un tel cas, *TA* n'est pas révélatrice.

C'est pourquoi l'amplitude totale ne suffit pas pour prédire l'emphase, et lorsque l'on se trouve entre 15,6 dB.s et 35,35 dB.s, d'autres critères entrent en ligne de compte.

B) Entre 15,6 et 21,5 dB.s d'amplitude totale, la fréquence fondamentale vient départager les classes d'emphase. Si l'écart entre la fréquence fondamentale moyenne du mot et celle de l'unité intonative entière est inférieur ou égal à 38,13%, le mot ne sera pas emphatique. Si cet écart est supérieur à 38,13%, le mot sera fortement emphatique. Ainsi, même si l'amplitude totale est assez basse mais que le mot se détache au niveau mélodique des autres mots de l'unité intonative, il sera très emphatique. Comme il s'agit d'un pourcentage positif, cela veut dire qu'il faut que le mot se détache vers le haut, c'est-à-dire qu'il doit être plus aigu en moyenne que le reste de l'UI.

Dans le segment P1.7S04, le mot « answer » a une amplitude totale de 20,11 dB.s, et l'écart entre la moyenne de la F0 sur le mot et la moyenne de la F0 de l'UI est de 72,6%. Ce mot est en effet emphatique pour 88,89% des auditeurs, c'est-à-dire qu'il est dans la classe 2 d'emphase.

De même dans le segment P1.8S06, le mot « in » a une amplitude totale assez faible (15,8 dB.s), ce qui n'est pas étonnant puisque c'est un mot court. C'est donc la fréquence fondamentale qui rend ce mot emphatique : l'écart entre la moyenne de la F0 sur le mot et sur l'UI est de 58,3%. Le mot est pointé emphatique par 83,33% des auditeurs.

C) Lorsque l'amplitude totale est comprise entre 21,5 et 35,35 dB.s, c'est d'abord à nouveau l'intensité qui départage les classes, puis la fréquence fondamentale entre en lice, mais il s'agit de la pente cette fois.

Ainsi si l'écart entre l'intensité maximale du mot et l'intensité moyenne de l'unité intonative est supérieur à 30,46%, la classe 2 est prédite. De ce fait, un mot peut être assez court, mais s'il a une forte intensité, il sera mis en classe 2.

Lorsque l'écart entre l'intensité maximale du mot et l'intensité moyenne de l'UI est inférieur ou égal à 30,46%, la fréquence fondamentale intervient, à travers la pente sans saut d'octave sur le mot (*MASWOJ_mot*) :

- si celle-ci est supérieure à 51,5 demi-tons/s (1957 Hz/s), le mot est emphatique. En d'autres termes, si sur un mot, la fréquence fondamentale subit une chute ou une montée (car il s'agit de la pente absolue) globale de plus de 50 dt/s par rapport à la moyenne de ce mot, le mot sera emphatique. La notion de mouvement mélodique est donc importante ici.
- si la pente est inférieure ou égale à 51,5 dt/s, c'est l'écart entre l'intensité maximale sur le mot et sur l'UI qui permet de faire des prédictions : s'il est faible, en dessous de -4%, la classe 0 est prédite ; s'il est au-dessus de -2,87%, la classe 1 (ou 2) est prédite, le mot sera emphatique¹²³ ; s'il se situe entre les deux valeurs (entre -4% et -2,87%), la pente est à nouveau déterminante. Si elle est inférieure à 28,47 dt, la classe 0 est prédite ; si elle est supérieure à 28,47 dt, la classe 2 est prédite.

On peut résumer cela en disant que l'amplitude totale est le paramètre le plus pertinent aux deux extrêmes. Si elle est faible, on a peu de chances d'avoir de l'emphase. Si elle est élevée, on a de grandes chances d'avoir de l'emphase. Cela est confirmé par la plupart des dix fichiers étudiés avec leur fichier de validation. Pour 9 d'entre eux, l'amplitude totale est le premier paramètre qui ressort.

Si la valeur de TA est moyenne, d'autres critères apparaissent : l'intensité et la F0. Si un mot est plus aigu ou a une intensité plus forte que les autres mots qui l'entourent dans l'unité intonative, il va être perçu emphatique. Si ce n'est pas le cas mais qu'en contrepartie le mot est prononcé avec un mouvement mélodique important, il va être également perçu emphatique. Cela est encore confirmé par le fichier *emph-0-8*, dont l'arbre est présenté sur la figure 36 ci-dessous :

¹²³ Nous trouvons des pourcentages négatifs ici car il s'agit de l'écart entre les deux valeurs maximales d'intensité, sur le mot et sur l'UI. Or nous avons vu en 3.3.2 que le mot le plus fort en intensité dans une unité intonative n'est pas toujours le mot le plus emphatique, mais souvent la tonique de tête.

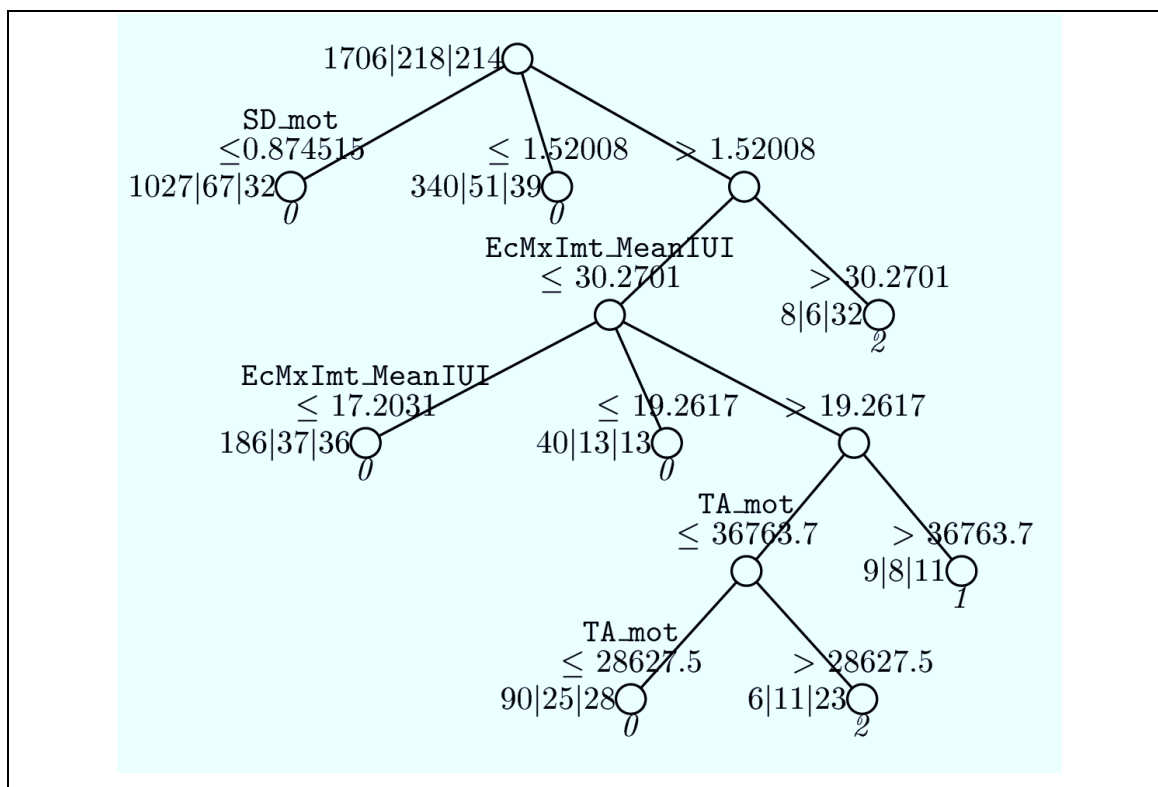


Figure 36 : arbre de régression pour le fichier emph-0-8, avec ISE.

Le paramètre SD_mot , c'est-à-dire l'écart-type sur le mot par rapport à la moyenne de la fréquence fondamentale, apparaît en tout premier. Si l'écart-type est faible (inférieur ou égal à 1,5 dt), la classe non emphatique est prédite. S'il est supérieur à 1,5 dt, l'intensité et l'amplitude totale entrent alors en jeu de façon comparable à ce que nous avons dit précédemment. Pour le corpus entier, nous avons vu que le paramètre pertinent pour la fréquence fondamentale est l'écart entre la moyenne du mot et la moyenne de l'UI. Dans le fichier emph-3-1¹²⁴, c'est l'écart entre la valeur maximale de l'UT et la moyenne de l'UI qui ressort. Dans emph-7-5, c'est l'écart entre la valeur maximale du mot et la moyenne de l'UI qui est plus pertinent. Dans emph-8-6, deux paramètres de fréquence fondamentale ressortent aussi : l'écart entre la valeur minimale du mot et la valeur moyenne de l'UI, et l'écart entre les moyennes du mot et de l'UI. Dans tous les cas, plus l'écart est élevé, plus le mot est emphatique. Le paramètre avec la valeur minimale est intéressant car ce sont des pourcentages positifs qui sont donnés de la façon suivante :

- si $EcMnmt_MeanFOUI \leq 38,41\%$, classe 0 ;

¹²⁴ Tous les arbres sont présentés en annexe 4b.

- si $EcMnmt_MeanF0UI > 38,41\%$, classe 2.

Cela veut dire que si l'écart entre la valeur minimale du mot et la valeur moyenne de l'UI est positif, donc si la valeur minimale du mot est plus élevée que la valeur moyenne de l'UI de plus de 38%, le mot sera très emphatique. Cela rejoint tout à fait nos considérations précédentes : lorsqu'un mot est plus aigu, en moyenne ou simplement sur sa valeur maximale, que les autres mots de l'UI, il est en général emphatique, comme c'est le cas dans le segment P1.1S05 sur le mot « so ». L'écart entre la valeur minimale de la F0 sur le mot et la moyenne de la F0 sur l'UI est de 46%, alors que l'amplitude totale est assez faible (20,3 dB.s), du fait qu'encore une fois il s'agit d'un mot court. Le mot est emphatique pour 88,89% des auditeurs.

De même, deux paramètres pour l'intensité semblent pertinents selon les différents fichiers : l'écart entre la valeur maximale sur le mot et la moyenne de l'UI, comme nous l'avons vu, mais également l'écart entre les valeurs maximales du mot et de l'UI (emph-1-9, emph-2-0, emph-6-4, emph-7-5) ; ce paramètre vient en général en deuxième dans les fichiers concernés. Encore une fois, plus l'écart est élevé, plus le mot a de chances d'être emphatique. Le segment P1.8S03, par exemple, confirme cette affirmation : le mot « debate » (emphatique pour 77,78% des auditeurs) n'a pas une amplitude totale élevée (20,18 dB.s), mais l'écart entre les valeurs d'intensité maximales du mot et de l'UI est de -0,05%, ce qui représente un faible écart, les écarts étant en grande majorité négatifs (la fourchette des valeurs pour ce paramètre va de 0,02 à -40,42%).

Nous avons vu qu'il n'est pas évident de trouver des constantes simples quant à l'expression de l'emphase. Cela est peut-être dû aux différents types de discours et nous avons voulu voir si les arbres pour chaque type de discours se révélaient très différents.

4.4.3 Types de discours¹²⁵

Pour le débat politique (P1), nous remarquons que TA_mot n'est pas le premier paramètre donné par CRUISE. On a tout d'abord l'écart-type sur le mot, qui concerne la

¹²⁵ Les arbres de classification pour chaque type de discours sont présentés en annexe 4c.

fréquence fondamentale. Si l'écart-type est supérieur à 5,5 dt, la classe 1 est prédite, avec également bon nombre de valeurs en classe 2. S'il est inférieur à 2 dt, la classe 0 est prédite. Le mouvement mélodique est donc déterminant ici. Entre les deux valeurs données ci-dessus, si l'écart entre l'intensité maximale du mot et celle de l'UI est supérieur à -2%, et que l'écart-type sur le mot est supérieur à 3,88 dt, la classe 2 est prédite. Ce sont donc l'intensité et le mouvement mélodique qui priment ici.

Ces résultats sont valables sur le corpus entier et nous avons enlevé 10% des données sur un seul fichier (P1-8-6.dat) pour tester les résultats : la matrice de confusion pour le fichier de validation est la suivante :

Classification Matrix for Test Data P1-7				
	Predicted Class			
	0	1	2	
Actual class	-----			
0	91	0	0	
1	7	0	0	
2	7	0	0	
Total obs in Test sample = 105, # correct = 91				

Les résultats sont donc très mauvais. Seule la classe 0 est prédite.

Nous pensons encore une fois que cela est peut-être dû aux stratégies différentes d'un locuteur à l'autre. Nous avons quatre locuteurs dans le débat politique, dont une seule femme, ce qui fausse peut-être les données, ou tout du moins les complique extrêmement.

En ce qui concerne la conversation, nous sommes confrontée au même problème : quatre locuteurs, deux hommes, deux femmes. Mais on a un équilibre ici et en effet l'essai avec le fichier de validation (fichier de données P2-8-6.dat et P2-7.dat pour valider) est nettement meilleur que pour le débat politique. Voici la matrice de confusion :

Classification Matrix for Test Data P2-7				
		Predicted Class		
		0	1	2
Actual class		-----		
	0	55	6	0
	1	7	4	0
	2	4	5	1
Total obs in Test sample = 82,		# correct = 60		

Les résultats sur les données globales donnent *TA* en premier lieu. Comme pour le corpus primaire entier, si *TA* est inférieure à 20 dB.s, la classe 0 est prédite. Si *TA* est supérieure à 40,7 dB.s, on se situe dans l'emphatique (classe 1 ou 2). Quand *TA* a une valeur moyenne, si l'écart entre l'intensité maximale du mot et l'intensité moyenne de l'UI est supérieur à 23%, on se situe aussi dans l'emphatique. Puis c'est à nouveau l'énergie qui apparaît, puis la moyenne de la fréquence fondamentale sur l'UT, encore l'énergie et enfin l'écart entre la moyenne de la F0 sur l'UT et sur l'UI. Nous parvenons à un arbre complexe, dans lequel l'énergie domine tout de même, revenant plusieurs fois. L'intensité est importante aussi et la F0 ne vient qu'après tout cela.

Les stratégies sont donc tout à fait différentes de celles du débat politique. Les prédictions étant meilleures ici, nous pensons que les stratégies des locuteurs sont peut-être moins variables que pour le débat politique, où l'on a affaire à des orateurs.

Enfin pour l'émission radio, l'arbre est beaucoup moins complexe. Peut-être encore une fois est-ce parce que l'on a moins de locuteurs (trois dont un parle très peu) et que ce sont tous les trois des femmes.

Nous avons extrait 10% des données comme pour les deux autres types de discours. La matrice de confusion sur le fichier de validation est la suivante :

Classification Matrix for Test Data P3-7				
		Predicted Class		
		0	1	2
Actual class		-----		
0		36	1	0
1		5	0	0
2		2	3	3
Total obs in Test sample = 50.		# correct = 39		

Nous estimons que ce sont de bons résultats, surtout sur un si petit nombre d'occurrences.

L'arbre sur le fichier global fait ressortir trois paramètres : l'énergie (TA), puis l'écart entre la $F0$ maximale sur le mot et la moyenne de l'UI quand TA a une valeur moyenne plutôt basse, et l'écart entre l'intensité maximale du mot et celle de l'UI quand TA a une valeur moyenne plutôt haute.

Comme pour le corpus primaire entier, si TA est faible (inférieure à 15,7 dB.s), on se situe dans le non emphatique. Si TA est élevée (supérieure à 36,5 dB.s) on se trouve dans l'emphatique (classe 1 ou 2).

Dans l'émission radio, c'est donc l'énergie et l'intensité ou la montée dans l'aigu qui sont déterminantes quant à la perception de l'emphase.

Il semble, au vu de ces analyses, que plus le nombre de locuteurs est élevé, plus on a de difficultés à parvenir à des prédictions correctes. La différence homme / femme est peut-être aussi un facteur important, bien que nous ayons tenté de l'atténuer en exprimant des écarts en pourcentage par exemple.

Nous nous sommes donc demandée plusieurs fois dans cette sous-section, mais également tout au long de ce travail, si l'on ne pouvait pas parler de stratégies différentes selon les locuteurs. C'est ce que nous allons tenter d'éclaircir.

4.4.4 Locuteurs

Pour chaque locuteur, nous avons fait les mêmes statistiques que ci-dessus (4.4.2 et 4.4.3), à la différence que comme nous ne disposons pas de données suffisamment nombreuses, nous n'en avons pas extrait 10% pour les tester. Nous ne disposons donc pas de fichier de validation pour nos locuteurs¹²⁶.

Intéressons-nous en premier lieu aux locuteurs du débat politique.

Pour la locutrice MM, le coefficient n°5 n'ayant donné aucun résultat, nous avons pris le coefficient n°4, avec seulement deux classes d'emphase (cf. 4.1.4).

¹²⁶ Les arbres de classification pour chaque locuteur sont présentés en annexe 4d.

Les quatre locuteurs sont partagés en deux groupes : chez DD et PM, l'amplitude totale (TA), c'est-à-dire l'énergie, puis la F_0 sont les paramètres les plus importants. Chez MM et JT, il s'agit de la F_0 en premier lieu puis de l'intensité.

Pour les deux premiers locuteurs, quand TA est inférieure à 22,6 dB.s (PM) ou 24,3 dB.s (DD), on se trouve en classe non emphatique.

Chez PM, lorsque TA est supérieure à 44,8 dB.s, on est en classe 2. Ainsi l'énergie est déterminante. Lorsque l'on se trouve entre les deux valeurs, la F_0 départage les classes : la pente sur le mot tout d'abord, puis l'écart-type, c'est-à-dire le mouvement mélodique et son amplitude. Plus ces deux paramètres sont élevés, plus le mot est emphatique.

Chez PM, lorsque la pente dépasse 48,3 dt/s (1631 Hz/s), le mot est prédit en classe 2 sans recours à l'écart-type.

Chez DD, l'arbre est différent, c'est la moyenne de la fréquence fondamentale sur l'UT qui est déterminante : si TA est supérieure à 24,3 dB.s et que $Mean_UT$ est supérieure à 157 Hertz, l'occurrence est prédite en classe 2. Dans tous les autres cas, la classe 0 est prédite. La classe 1 ne sort pas dans les prédictions. Chez ce locuteur c'est donc la hauteur de la voix davantage que le mouvement mélodique qui importe pour l'emphase.

Chez JT et MM, le premier paramètre donné est l'écart-type sur le mot.

Chez JT, si $SD_mot < 2dt$ (112 Hz), la classe 0 est prédite ; si $SD_mot > 2,68 dt$ (116,7 Hz), la classe 2 est prédite ; entre ces deux valeurs, l'intensité entre en jeu, mais sur peu d'occurrences , et cela ne nous paraît pas suffisamment révélateur. L'écart-type est donc le paramètre le plus important, c'est-à-dire la déviation par rapport à la moyenne, donc le mouvement mélodique.

Chez MM, si $SD_mot \leq 1,13 dt$ (106,7 Hz), la classe 0 est prédite. Dans les autres cas, l'intensité et la F_0 interviennent. Pour l'intensité, c'est l'écart entre l'intensité maximale du mot et celle de l'UI qui est déterminant : s'il est inférieur à -2,6%, on se situe dans le non emphatique. Ensuite si l'écart entre la F_0 maximale de l'UT et la moyenne de l'UI est supérieur à 43%, on est dans l'emphatique. En résumé, l'emphase est rendue par MM grâce à un mouvement mélodique qui part de l'aigu, ainsi qu'une hausse de l'intensité par rapport à la moyenne.

Après l'analyse des données de ces quatre locuteurs, nous comprenons mieux pourquoi les prédictions sur le fichier de validation du débat politique sont si mauvaises. Chacun a une façon différente d'exprimer l'emphase.

En ce qui concerne les locuteurs de la conversation, nous avons rencontré le même problème avec DL qu'avec MM ci-dessus : nous ne sommes parvenue à aucun résultat avec le coefficient d'emphase n°5. Nous avons donc pris le coefficient n°4 avec seulement deux classes (emphatique et non emphatique). Quant au locuteur JC, ni le coefficient 5, ni le coefficient 4 n'ont donné de résultat. De façon tout à fait étonnante, le coefficient n°3 (avec 4 classes d'emphase) nous a permis d'obtenir un arbre. Cela est d'autant plus surprenant que JC fait partie des locuteurs pour lesquels nous avons peu d'occurrences (102). L'arbre obtenu comprend plusieurs arborescences qui sont, du fait du petit nombre d'occurrences, basées sur très peu d'exemples. Nous ne sommes donc pas satisfaite de ces résultats et n'en ferons pas beaucoup cas.

On remarque que pour tous les locuteurs de la conversation, le premier paramètre est TA_mot (l'énergie).

Chez VB, c'est même le seul paramètre qui ressort :

- si $TA \leq 20,5$ dB.s, la classe 0 est prédite ;
- si TA est comprise entre 20,5 et 39,3 dB.s, c'est la classe 1 ;
- enfin si $TA > 39,3$ dB.s, la classe 2 est prédite.

Le cas de VB est donc très simple. La différence entre les classes d'emphase dépend principalement et en général de l'énergie émise pour chaque mot.

Chez tous les locuteurs, si TA est plutôt faible, c'est la classe non emphatique qui est prédite. Si TA est élevée, c'est la classe la plus emphatique qui sort (sauf pour JC). Il est alors intéressant de voir quelle est la valeur limite de TA pour chaque locuteur.

La limite inférieure est de 20,5 (VB), 20,3 (CB), 15 (JC), et 12 dB.s (DL). Pour la classe la plus emphatique, TA doit être supérieure à 39,3 (VB), 39,8 (CB) et 28 dB.s (DL). Il semble y avoir des valeurs très proches, mais différentes selon qu'il s'agit d'un homme ou d'une femme¹²⁷.

Pour DL et CB, lorsque TA est comprise entre les deux valeurs données ci-dessus, d'autres paramètres départagent les classes : l'intensité pour DL et plutôt la fréquence fondamentale pour CB. Ceci est intéressant car nous avons vu lors de l'analyse acoustique (chapitre 3) que DL prononçait souvent ses énoncés de façon très plate. La fréquence fondamentale n'est donc pas révélatrice chez lui, c'est la hausse ou la baisse d'intensité qui fait la différence. Le paramètre pertinent dans son cas est l'écart entre

¹²⁷ Nous rappelons que VB et CB sont des femmes, JC et DL des hommes.

l'intensité maximale du mot ou de l'UT et l'intensité maximale de l'UI. L'écart est exprimé en pourcentage et ce sont ici des pourcentages négatifs, ce qui veut dire que l'intensité maximale de l'UI est plus élevée que celle du mot ou de l'UT. De ce fait, moins l'écart est élevé, plus le mot est emphatique. Cela va dans le sens de ce que nous avons déjà vu plus haut pour le corpus primaire.

Chez CB, les paramètres moyenne de la fréquence fondamentale de l'UI, amplitude totale à nouveau, écart-type sur le mot et écart entre la pente sur l'UT et sur l'UI ressortent. De la même façon, plus les valeurs sont élevées, plus l'occurrence a de chances d'être emphatique. Ce sont donc plutôt la hauteur et le mouvement mélodique qui comptent chez CB.

Prenons enfin les deux locutrices de l'émission radio (nous avons éliminé HN, car nous avons trop peu d'occurrences pour cette locutrice).

Comme chez VB, l'emphase chez YB semble dépendre principalement de TA :

- si $TA \leq 21,5$ dB.s, la classe 0 est prédite ;
- si TA est comprise entre 21,5 et 31,2 dB.s, c'est la classe 1 ;
- si $TA > 31,2$ dB.s, c'est la classe 2.

On remarque que les valeurs sont assez proches de celles de VB.

Chez PS, TA ne vient qu'après l'écart entre l'intensité maximale du mot et l'intensité moyenne de l'UI. Si cet écart dépasse 31%, on se trouve dans la classe la plus emphatique. Si l'écart est inférieur à 31%, alors TA , puis à nouveau l'intensité, mais cette fois l'écart entre les valeurs maximales du mot et de l'UI, entrent en jeu. C'est donc uniquement l'intensité et l'énergie qui sont révélatrices pour l'emphase chez cette locutrice, et non la $F0$.

Comme nous l'avions supputé, l'emphase est rendue de différentes façons selon les locuteurs. Chez certains, l'énergie est déterminante, chez d'autres c'est la fréquence fondamentale : le mouvement mélodique souvent, ou bien la hauteur tout simplement.

Même si l'on retrouve des constantes, nous pensons que cela constitue un obstacle à la bonne prédiction de l'emphase, et cela explique pourquoi CRUISE ne parvient pas à des prédictions très élevées en ce qui concerne la classe la plus emphatique pour le corpus entier et pour le débat politique en particulier.

4.5 *Résumé et conclusion*

L'analyse statistique entreprise grâce à un logiciel tout récent et extrêmement performant devait nous permettre d'extraire les paramètres les plus révélateurs de l'emphase sur 2375 occurrences et 11 locuteurs dans 3 types de discours différents.

Nous avons opté pour une notion dichotomique de l'emphase avec deux degrés dans l'emphatique.

La diversité du corpus et le nombre de locuteurs ont constitué une difficulté majeure en ce qui concerne les prédictions. Cela signifie qu'il existe plusieurs moyens d'exprimer l'emphase en anglais et nous avons vu que ces moyens variaient considérablement d'un locuteur à l'autre.

Grâce à la performance du logiciel CRUISE, nous pouvons néanmoins dégager des constantes et affirmer que l'énergie est un paramètre fondamental dans la perception de l'emphase. Cela signifie que si un mot est allongé et prononcé avec une intensité plus forte que sur les mots adjacents, il sera généralement perçu emphatique. Si au contraire l'énergie dépensée est faible, le mot sera perçu non emphatique.

Lorsque l'énergie n'est pas déterminante, alors la hauteur mélodique le devient et cela corrobore les hypothèses émises dans le chapitre 3 : le mouvement mélodique sur le mot est important. S'il est ample et qu'il part de plus haut que pour les mots adjacents, le mot sera perçu emphatique.

Enfin l'intensité est déterminante aussi. Plus l'intensité est forte par rapport aux mots adjacents, plus le mot sera perçu emphatique.

N'oublions pas de nuancer ces affirmations. Nous avons vu que ces règles sont générales et ne marchent pas à 100%. En effet, chaque locuteur semble avoir ses habitudes et ses façons propres d'exprimer l'emphase, ses stratégies personnelles.

Cette étude statistique confirme que l'emphase est un phénomène complexe mais que l'on peut tout de même énoncer certains principes généraux.

5 Chapitre 5

Etude perceptive sur signaux manipulés

Nous avons pensé qu'il serait bon de ne pas s'arrêter à la machine et aux résultats qu'elle nous donne, et qu'il pourrait être intéressant de comparer les résultats de l'analyse objective avec des résultats basés sur la perception humaine. Ainsi nous avons réalisé, en parallèle à l'étude statistique, une étude que nous appelons « étude perceptive sur signaux manipulés », afin de tester nos hypothèses de départ, à savoir que les paramètres essentiels dans la perception de l'emphase sont la fréquence fondamentale, la durée segmentale, l'intensité et peut-être les pauses. Un autre paramètre est le paramètre contextuel, pragmatique, qu'il nous a été impossible d'insérer dans l'analyse statistique. Cette étude perceptive nous permet d'explorer ce paramètre-là en plus, comme nous allons le montrer.

L'idée est de partir de stimulus resynthétisés, manipulés et non manipulés, et de demander à des anglophones, à des « natifs », de pointer les passages emphatiques, comme pour l'étude perceptive qui a servi de point de départ à tout ce travail¹²⁸. Nous n'avons pas sollicité les mêmes auditeurs que pour les sondages décrits au chapitre 2, afin qu'ils arrivent avec une oreille vierge.

Il existe quelques exemples d'études similaires dans la littérature et nous nous sommes inspirée principalement de Pagel *et al.* (1996) pour l'idée de départ et de Ramus & Mehler (1999) pour la réalisation de cette étude.

¹²⁸ Cf. 2.2.

5.1 Test de perception

5.1.1 La resynthèse vocale : principes généraux

La resynthèse vocale est une technique qui a été utilisée pour l'étude de la prosodie à l'IPO d'Eindhoven, et pour la délexicalisation par Pagel *et al.* (1996). Elle consiste en deux phases distinctes :

- mesurer toutes les composantes acoustiques pertinentes du signal ;
- utiliser ces mesures ainsi qu'un algorithme approprié pour resynthétiser la parole.

L'intérêt de la resynthèse dans notre étude est de manipuler les signaux en modifiant justement certaines des mesures afin de voir quel paramètre est le plus pertinent dans la perception de l'emphase.

Des segments du corpus primaire ont donc été choisis, il s'agit de phrases ou d'unités intonatives qui ont été données très emphatiques par la plupart des auditeurs et qui nous semblaient intéressantes. Ces segments ont été resynthétisés, puis manipulés, selon quatre critères : le sémantisme, la fréquence fondamentale, la durée segmentale, les pauses. Nous avons choisi de ne pas mélanger ces quatre paramètres. L'intensité n'a pu être intégrée à l'étude car l'algorithme de resynthétisation ne permet pas de le faire.

5.1.2 Phrases-sources

Comme nous l'avons expliqué dans les principes généraux, nous avons choisi pour phrases-sources des phrases (ou unités intonatives) dans lesquelles au moins un mot a été marqué comme emphatique par la plupart des auditeurs du test de perception de départ. Nous avons sélectionné des phrases contenant des pauses, afin de pouvoir les enlever, des phrases qui comportaient des allongements de phonèmes, et des phrases avec des mouvements mélodiques amples ainsi que des phrases avec très peu de mouvement mélodique. Nous avons également pensé au côté sémantique puisque nous

procédons à la délexicalisation de toutes les phrases. Douze segments ont été choisis, afin que le test ne soit pas trop long pour les auditeurs car en effet, pour chaque segment, nous aurons cinq phrases resynthétisées, cela donnant un total de soixante stimulus.

Voici donc les passages sélectionnés¹²⁹ :

P1.1S01 (↵ exs.58) : DD *This is a = kerfuffle today over whether the Chancellor had threatened to resign*

Le sémantisme est important dans cette phrase : le mot « kerfuffle » n'est pas courant, ni anodin, il n'est donc pas neutre et peut-être est-ce une des raisons pour lesquelles il est emphatique (à 83,33%). Il y a aussi une pause avant « kerfuffle » et un grand mouvement mélodique sur ce même mot. Quant à « resign », qui est également emphatique à 83,33%, il est en chute et la dernière syllabe subit un allongement.

P1.3S01 (↵ exs.59) : JT : *looking at it I think we're certainly going to have a better chance of snow = on Christmas Day*

Les mots « snow » et « Christmas » sont emphatiques à 94,44% et 77,77% respectivement. On a une pause ici mais qui se trouve au changement d'UI. Elle n'est donc sûrement pas révélatrice en ce qui concerne l'emphase. Contrairement à la phrase précédente, on constate un très faible mouvement mélodique, de type ton creusé, sur « snow ». La courbe est plate avant « snow ». Il sera intéressant de voir si ce faible mouvement suffit à faire percevoir une emphase. Si l'on se base sur des valeurs de durée moyenne, le mot « snow » est plus long qu'à la normale.

P1.4S04 (↵ exs.60) : PM : *there's a perfectly sensible policy which the Conservative = cabinet the Conservative government and the Conservative Party in the House of Commons*

C'est le mot « cabinet » qui a été donné le plus emphatique dans ce passage (94,44%), mais on trouve aussi « government » (88,88%) et « perfectly » (77,77%). Or on a une pause juste avant « cabinet », qu'il sera intéressant d'enlever. D'autre part le

¹²⁹ Les pauses sont symbolisées par le signe égal, et la ponctuation a été volontairement omise.

sémantisme est peut-être important, ceci dû à la répétition de « Conservative ». La F0 est en montée sur la fin de l'unité intonative, ainsi que sur « cabinet ».

P1.4S05 (☞ exs.61) : JT : *and that is that we're going to wait to see = how this argument develops = on a single currency*

Ici le mot « currency » est le plus emphatique (100%) et ce qui précède est de type ondulatoire (« sing song ») avec « wait » et « see » tous deux donnés à 77,77%. On se demande donc si cette mélodie ondulante joue un rôle dans la perception de l'emphase. De plus, le mot « currency » est en grande chute. On note la présence de deux pauses.

P1.8S05 (☞ exs.62) : MM : *what we want to do is actually get in there and represent = our country's interests*

Les mouvements mélodiques sur « in » et sur « represent », qui sont les mots les plus emphatiques avec 88,88% pour les deux, sont des chutes assez amples, et on a aussi une pause juste après « represent ». Le schéma intonatif de la deuxième unité intonative (à partir de « and ») est de type « dive », c'est-à-dire que l'on a une chute sur le noyau qu'est « represent », ensuite toute la queue de l'énoncé est en statique grave, sauf la dernière syllabe qui remonte. Ce schéma est habituellement utilisé en contextes non neutres justement.

P1.9S04 (☞ exs.63) : PM : *Now it's slow progress but they are making progress*

Cet énoncé ne comporte pas de pause. Le passage est intéressant car sont mis en noyaux des mots inattendus à cette place du fait de leur catégorie grammaticale puisqu'il s'agit de « slow » et « are ». Le premier est considéré emphatique pour 94,44% des auditeurs et le second pour 100% d'entre eux. Les mots sont-ils donc déterminants ici ? Ou est-ce la fréquence fondamentale, car les deux mots comportent un mouvement mélodique ? Ou est-ce plutôt l'allongement des mots ? Ou est-ce une combinaison de tout cela ?

P2.1S02 (☞ exs.64) : JC : *Did you do that = last year as well ?*

« Last » a été donné emphatique (87,5%). On a certainement un allongement, mais aussi une pause avant. Le mouvement mélodique n'est pas particulièrement emphatique (il n'est pas ample). Il est montant.

P2.4S01 (↵ exs.65) : CB : *and it was absolutely = freezing*

« Absolutely » et « freezing » sont tous deux emphatiques (75% et 93,75 % respectivement). On a une pause entre les deux mots. Le mouvement mélodique n'est pas très important. Ce passage est intéressant pour le sémantisme car les adverbes de ce type sont souvent mis en relief, et « freezing » n'est pas neutre non plus, mais exagéré bien sûr. Le mot emphatique est en position finale cette fois.

P2.4S03 (↵ exs.66) : JC : *and the pilot said well there's a ten-degree difference between Marseille and and and London*

Tout l'énoncé est prononcé de façon très plate, et « ten » est donné emphatique (93,75%). Cela est donc intéressant et peut même paraître étonnant. Une toute petite montée de F0 (sur « ten ») suffit-elle à rendre un mot emphatique ou est-ce dû à autre chose ? On n'a pas de pause dans cette phrase. Nous verrons ce que cela donne si l'on en rajoute une.

P2.7S04 (↵ exs.67) : VB : *I think people who learn a language = just by picking it up you know as adults = then they probably do pick up all sorts of mistakes*

Le mot « just » est emphatique (93,75%) dans la première unité intonative (jusque « adults »). On a une pause avant. La F0 est-elle déterminante ? Il y a en effet un mouvement mélodique assez ample. Ou est-ce le sémantisme du mot ? Ou encore la consonne initiale, qui est très appuyée par la locutrice ?

Dans cette phrase, la deuxième unité intonative comporte un « do » que l'on appelle traditionnellement le « do » emphatique. Or ici, « do » n'a pas de mouvement mélodique particulier, ni d'allongement, ni de pause avant ou après. Ce passage a donc son importance au niveau du sémantisme, de la syntaxe. Mais l'emphase peut-elle être due à autre chose qu'à la syntaxe ? Nous avons choisi délibérément un des « do » emphatiques donnés les moins emphatiques par nos auditeurs (68,75% tout de même et 43,75% pour le degré d'emphase). Le degré d'emphase nous semble révélateur ici, car

la syntaxe force certainement les auditeurs à considérer le mot emphatique. Mais peut-être est-ce un autre critère qui leur permet de déterminer un degré d'emphase¹³⁰.

P3.1S04 (☞ exs.68) : YB : *It's not = even been discussed till this report came out properly*

On trouve ici un mot emphatique (« even », 93,75%) précédé d'une pause et dont la voyelle initiale est probablement allongée. Cette fois le mot se trouve plutôt vers le début de la phrase. Au niveau de la fréquence fondamentale on note une chute sur « even ».

P3.2S03 (☞ exs.69) : YB : *or would you say domestic violence = is a private issue*

La pause se situe après le mot emphatique (« violence », 100%) dans cette phrase. Le sémantisme a son importance en contexte : il est question de violence domestique depuis le début de l'émission et cette forme de violence n'est pas en opposition avec une autre forme de violence. Nous sommes hors contexte ici puisque les phrases sont données seules, contrairement à ce qui a été fait dans le test de perception de départ où les auditeurs entendaient un long morceau de contexte. Nous verrons si c'est alors plutôt « domestic » qui est donné emphatique hors contexte¹³¹. La durée joue peut-être un rôle mais c'est sûrement la F0 qui est plus pertinente ici car on a une grande chute sur « violence », et le schéma global est de type « High Jump », c'est-à-dire que l'on a une montée préparatoire dans le corps de l'énoncé jusqu'à la grande chute qui se situe sur le noyau et la queue est en statique grave.

Ainsi nous avons différents types de phrases, prononcés par des locuteurs différents. Nous avons choisi des locuteurs différents car nous avons vu que les locuteurs peuvent avoir des stratégies différentes et notre test ne sera donc pas faussé par cela.

¹³⁰ Nous avons cinq autres « do » emphatiques dans le corpus primaire, dont nous donnons le pourcentage d'emphase suivi du pourcentage de degré d'emphase : P2.1S04, did (81,25/66,66), P2.6S06, do (100/91,66), P2.6S07 (75/56,25), P3.1S05, do (93,7/5/85,41), P3.3S05, does (37,5/29,16).

¹³¹ Il faut préciser que « domestic » atteint quand même le score de 62,5%.

5.1.3 MBROLA

Pour resynthétiser ces phrases, nous avons utilisé le logiciel MBROLA (Dutoit *et al.*, 1996 ; Dutoit, 1997). MBROLA est un synthétiseur de parole basé sur la concaténation de diphtonges. Le projet a été mis en place par le laboratoire TCTS de la Faculté Polytechnique de Mons en Belgique et le logiciel est distribué gratuitement pour des applications non commerciales et non militaires ([http: cf. références bibliographiques](http://cf.références bibliographiques)). MBROLA prend comme entrée une liste de diphtonges associée à de l'information prosodique (durée et hauteur) et sort des échantillons de parole à 16 bits.

MBROLA fonctionne avec des fichiers *.pho*, qui contiennent les phonèmes en écriture SAMPA, leur durée et éventuellement la hauteur mélodique. On peut rentrer plusieurs valeurs de F0 pour un même phonème, il suffit de préciser en pourcentage à quel moment du phonème la valeur de F0 prend effet. On peut aussi ne rentrer aucune valeur de F0 pour un phonème et une interpolation linéaire a alors lieu entre les deux valeurs qui l'entourent avec une valeur toutes les 10 millisecondes. On peut ainsi obtenir une parole synthétique d'excellente qualité, très proche de la parole naturelle.

Afin d'utiliser MBROLA, nous avons segmenté les douze passages choisis en phonèmes, avec le logiciel PRAAT, de la même façon que nous avons fait la segmentation en mots, c'est-à-dire grâce à l'oscillogramme, au spectrogramme, à la courbe de F0 et à l'oreille¹³². Nous avons simplement rajouté un étage (« tier ») nommé phonèmes, comme l'illustre la figure 37¹³³. Les phonèmes sont donnés en alphabet SAMPA¹³⁴.

¹³² Cf. 2.3.2.2.

¹³³ Les fichiers TextGrid avec l'étage des phonèmes pour les douze phrases-sources sont disponibles sur le CD ROM joint : *Etiquetage : TextGrid phonèmes*.

¹³⁴ Cf. annexe 1 ou CD ROM : *IPA/SAMPA*.

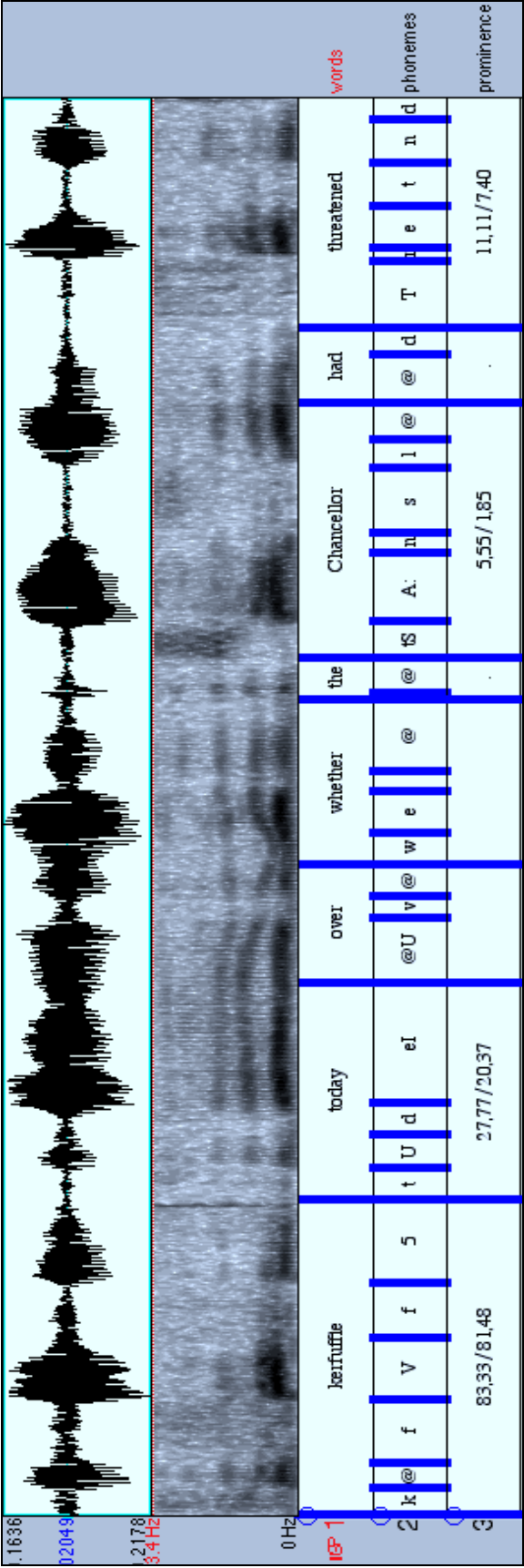


Figure 37 : P1.IS01, fichier son + TextGrid avec phonèmes en alphabet SAMPA.

Grâce à PRAAT, nous avons obtenu à partir de notre segmentation manuelle la durée de tous les phonèmes¹³⁵, que nous rentrons juste après les phonèmes dans le fichier *.pho*. Nous avons mesuré manuellement la fréquence fondamentale sur toutes les voyelles de nos phrases-sources, ainsi que sur certaines consonnes. Nous avons, en règle générale, pris une seule valeur de F0 quand les mots ne présentaient pas de mouvement mélodique particulier, et nous avons rentré jusqu'à quatre valeurs quand cela nous semblait nécessaire. Dans le segment P1.3S01 (☞ **exs.59**) par exemple, nous avons déjà dit que « snow » et « day » sont prononcés avec un ton creusé. Dans ces cas de tons complexes, nous avons rentré quatre valeurs¹³⁶ :

s									
n	50	123							
@U	10	126	34	111	62	93	93	95	
d	50	110							
eI	14	104	45	82	83	120	98	110	

La base de diphones pour MBROLA anglais britannique a été enregistrée par un locuteur masculin. Cela ne pose donc pas de problème pour les valeurs de F0 de nos locuteurs masculins mais est gênant pour les locutrices dont les valeurs de F0 sont plus élevées.

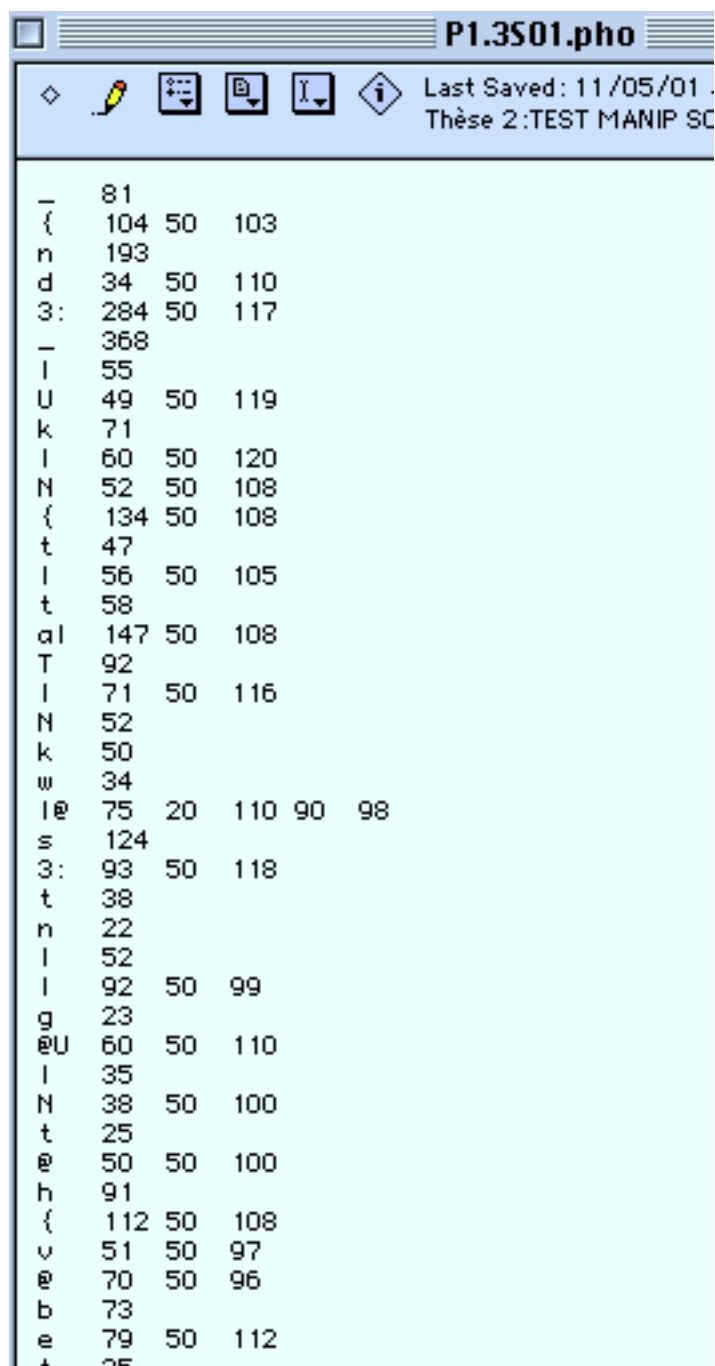
On constate généralement que la moyenne basse pour une voix d'homme se situe entre 90 et 135 Hz, alors que c'est plutôt de 120 à 180 Hz pour une femme, ce qui fait un rapport de 1,3. Pour les passages prononcés par des femmes, nous avons donc divisé toutes les valeurs de F0 par 1,3 avant de les rentrer dans les fichiers *.pho*. Une correction de ce type revient à une normalisation sur une échelle logarithmique, mieux adaptée à la perception de l'oreille humaine qu'une échelle linéaire.

La figure 38 montre un exemple de fichier *.pho*¹³⁷ (la première colonne comporte les phonèmes en alphabet SAMPA, la seconde colonne les durées, et les autres colonnes les paires exprimant la F0) :

¹³⁵ Nous avons utilisé la fonction « Table of Real » de PRAAT, qui donne les durées de toutes les étiquettes du TextGrid.

¹³⁶ Les valeurs marchent par paires, le premier chiffre de la paire étant le pourcentage (50 équivaut à la moitié du mot par exemple, 75 aux trois-quarts, etc...) et le second chiffre la valeur de F0 en Hertz lui correspondant.

¹³⁷ Les pauses dans MBROLA sont symbolisées par un tiret.



Symbol	Value 1	Value 2	Value 3	Value 4	Value 5
-	81				
{	104	50	103		
n	193				
d	34	50	110		
3:	284	50	117		
-	368				
l	55				
U	49	50	119		
k	71				
l	60	50	120		
N	52	50	108		
{	134	50	108		
t	47				
l	56	50	105		
t	58				
al	147	50	108		
T	92				
l	71	50	116		
N	52				
k	50				
w	34				
le	75	20	110	90	98
s	124				
3:	93	50	118		
t	38				
n	22				
l	52				
l	92	50	99		
g	23				
@U	60	50	110		
l	35				
N	38	50	100		
t	25				
@	50	50	100		
h	91				
{	112	50	108		
v	51	50	97		
@	70	50	96		
b	73				
e	79	50	112		
t	35				

Figure 38 : P1.3501.pho

Ainsi nous avons obtenu douze phrases resynthétisées aussi proches que possible des phrases originales¹³⁸. Il était nécessaire de fabriquer ces phrases afin de constituer un point de repère sur lequel se fonder lors de l'interprétation des résultats. Nous aurions

¹³⁸ On peut entendre tous les stimulus sonores, qui sont disponibles sur le CD ROM : *Synthèse voc : Fichiers son : Normal*.

pu nous baser sur la phrase originale, mais avec le risque de trouver un trop grand écart entre celle-ci et toutes les autres. En effet, s'il est vrai que MBROLA est un synthétiseur de parole performant, on perçoit tout de même sur certains mots ou certains enchaînements qu'il s'agit de parole artificielle. C'est la raison pour laquelle nous ne prendrons pas pour point de comparaison la phrase originale, mais la phrase resynthétisée sans manipulation, de façon à ce que nos résultats ne soient pas faussés.

Il sera en outre intéressant de comparer les résultats obtenus sur la phrase originale et ceux obtenus à partir de la phrase resynthétisée sans manipulation. En effet, la phrase est resynthétisée sur la base de deux paramètres uniquement (F0 et durée), et d'autre part, les phrases synthétiques sont données hors contexte ici, alors que l'on avait donné le contexte lors des premiers sondages sur les phrases originales.

Une fois les stimulus synthétiques non modifiés créés, nous avons fabriqué des phrases dans lesquelles nous avons changé certaines données.

5.1.4 Manipulation des paramètres acoustiques et prosodiques

5.1.4.1 Durées segmentales

Nous avons resynthétisé les mêmes phrases en changeant les durées segmentales pour chaque phonème. Plusieurs options se présentaient :

- allonger ou raccourcir certains phonèmes, judicieusement choisis en fonction du degré d'emphase ;
- prendre une valeur moyenne pour chaque phonème afin de supprimer toute variation de durée de type emphatique.

C'est pour la deuxième solution que nous avons opté, celle-ci nous semblant moins aléatoire et moins subjective que la première.

Nous avons utilisé les données de Campbell (1992b), qui a obtenu des valeurs moyennes pour chaque phonème de l'anglais (cf. 3.2.2).

Nous avons ainsi obtenu douze phrases, étiquetées P...S..._D¹³⁹ : le premier fichier, par exemple, est nommé P1.1S01_D.

5.1.4.2 Fréquence fondamentale

Comme nous l'avons dit ci-dessus (5.1.3), MBROLA interpole linéairement entre deux valeurs consécutives de F0. On n'est donc pas obligé de rentrer une valeur pour chaque phonème. On peut même ne rentrer que deux valeurs, une sur le premier mot et une deuxième sur le dernier mot. C'est ce que nous avons fait afin que la courbe de F0 soit plate, que l'on ait un effet « monotone ». Cependant, pour que la phrase ne sonne pas trop artificielle, nous avons créé un effet de déclinaison en rentrant 135 Hz pour le premier mot et 90 Hz pour le dernier, valeurs moyennes basses pour une voix d'homme. Nous avons obtenu douze phrases étiquetées P...S..._F¹⁴⁰ : P1.1S01_F, par exemple.

5.1.4.3 Pauses

La plupart de nos phrases-sources comportent des pauses. Nous les avons éliminées, en gardant les valeurs originales de durée et de fréquence fondamentale.

Nous avons rajouté des pauses de façon complètement arbitraire quand il n'y en avait pas (nous avons mis des pauses de 350 ou 400 ms selon les cas, en fonction du locuteur).

C'est le cas pour les passages :

P1.3S01 : ajout d'une pause de 400 ms devant « better » ;

P1.9S04 : ajout de pauses de 350 ms devant « progress » (2 fois) ;

P2.4S03 : ajout d'une pause de 400 ms devant « ten » ;

P2.7S04 : ajout d'une pause de 300 ms devant « do » ;

Les douze passages sont étiquetés P...S..._P¹⁴¹.

¹³⁹ Cf. CD ROM : *Synthèse voc : Fichiers .pho : Durée* pour accéder aux fichiers .pho et *Synthèse voc : Fichiers son : Durée*, pour entendre les stimulus.

¹⁴⁰ Cf. CD ROM : *Synthèse voc : Fichiers .pho : F0* pour accéder aux fichiers .pho et *Synthèse voc : Fichiers son : F0*, pour entendre les stimulus.

¹⁴¹ Cf. CD ROM : *Synthèse voc : Fichiers .pho : Pauses* pour accéder aux fichiers .pho et *Synthèse voc : Fichiers son : Pauses*, pour entendre les stimulus.

5.1.5 Délexicalisation

Enfin nous avons pu tester le critère du sémantisme, grâce à la délexicalisation. Il s'agit de garder les critères acoustiques et prosodiques (la fréquence fondamentale, les durées et les pauses ici) en changeant les phonèmes afin d'obtenir des mots inconnus. Ainsi les critères du sens des mots et du contexte de la phrase disparaissent.

On trouve également ce genre de transformations dans la littérature et là encore plusieurs options sont possibles.

Gussenhoven & Rietveld (1988) ont utilisé « da-da ». Sluitjer & van Heuven (1993) ont utilisé « nana », mais pour un seul mot au milieu d'une phrase. D'autres encore ont choisi « mama ».

Il semble y avoir un consensus autour de la voyelle /a/, car c'est effectivement la voyelle la plus ouverte et la plus longue dans beaucoup de langues. En ce qui concerne les consonnes, différentes approches existent et Ramus & Mehler (1999 : 514) présentent trois possibilités (qu'ils ont toutes utilisées d'ailleurs) :

- ils ont nommé la première transformation « saltanaj » : toutes les voyelles sont remplacées par /a/, toutes les fricatives par /s/, les plosives par /t/, les liquides par /l/, les nasales par /n/ et les glissantes par /j/. Ces phonèmes ont été choisis car ce sont les plus universaux, tous présentant une articulation alvéolaire ou coronale.
- le deuxième type de transformation est appelé « sasasa », car toutes les voyelles sont transformées en /a/ et toutes les consonnes en /s/. La consonne /s/ est sélectionnée pour son caractère continu, qui permet à une succession de consonnes de ne sonner que comme un seul /s/ long ;
- enfin la troisième transformation consiste à remplacer tous les phonèmes par /a/ : « aaaa ». De même il ne s'agit pas d'une suite de /a/, mais d'un seul /a/, variant continuellement en fonction de la hauteur mélodique et de la durée (à part éventuellement dans le cas de pauses).

C'est la première solution qui nous semble la plus intéressante pour l'étude que nous voulons mener, car en effet cette option permet de suivre et d'annoter un texte, ce qui est essentiel pour nous. Il faut préciser en effet que Ramus & Mehler (1999) travaillent sur plusieurs langues, et non pas seulement l'anglais et plus particulièrement sur le rythme, ce qui n'est pas notre cas.

Nous avons donc opté pour « saltanaj » mais en le modifiant un peu : nous avons remplacé les voyelles par /a/, les plosives sourdes par /t/, les plosives sonores par /d/, les nasales par /n/, les fricatives sourdes par /s/, les fricatives sonores par /z/, /w/ et /j/ par /j/ et enfin /l/ et /r/ par /r/ plutôt que par /l/. On notera que /r/ est, comme /l/, une consonne avec une articulation coronale.

Cette formule appelée « jastradan » est mieux adaptée à l'anglais en raison des groupes consonantiques. En anglais, on ne trouve pas /tl/ à l'initiale d'un mot par exemple. Avec la formule « saltanaj », tous les mots commençant par une plosive suivie d'un /r/ se transforment en mots commençant par /tl/. Avec la formule « jastradan », /pr/, /tr/ et /kr/ se transforment en /tr/, et /br/, /dr/ et /gr/ deviennent /dr/.

Un script en langage Perl, *mapPhones*¹⁴², nous a permis de convertir les phonèmes de tous nos fichiers *.pho* automatiquement. Ainsi nous avons encore obtenu douze phrases délexicalisées avec les durées et les hauteurs mélodiques originales.

Suivant l'exemple de Ramus & Mehler (1999) ces phrases ont été resynthétisées à partir de la version française de MBROLA (voix d'homme), afin de rendre plus crédible l'idée qu'il s'agissait d'une autre langue. Le nombre plus important de consonnes dans la formule « jastradan » comparée à « saltanaj » contribuait également à cet effet.

Ces phrases sont étiquetées P...S..._B¹⁴³.

5.1.6 Mise en place du test

Nous avons procédé comme pour le premier test de perception de cette étude : nous avons demandé à des auditeurs¹⁴⁴ d'écouter les phrases resynthétisées et de souligner les passages, mots ou syllabes emphatiques sur un papier sur lequel les phrases sont transcrites orthographiquement sans ponctuation¹⁴⁵. Les phrases sont données deux fois chacune.

¹⁴² Cf. annexe 5a ou CD ROM : *Scripts : PERL : mapPhones*.

¹⁴³ Cf. CD ROM : *Synthèse voc : Fichiers .pho : jastradan* pour accéder aux fichiers *.pho* et *Synthèse voc : Fichiers son : jastradan*, pour entendre les stimulus.

¹⁴⁴ Cf. annexe 5b ou CD ROM joint : *Synthèse voc : Test perceptif : Auditeurs*.

¹⁴⁵ Cf. annexe 5c ou CD ROM : *Synthèse voc : Test perceptif : Instructions*.

Nous avons 60 phrases et cela nous semblait beaucoup pour un seul auditeur. Nous avons alors décidé de séparer nos phrases non délexicalisées en deux groupes équivalents (c'est-à-dire ayant le même nombre de phrases avec la F0 modifiée, les pauses, les durées modifiées et les phrases « normales ») afin de rendre le test plus court.

Dans le test, nous avons d'abord présenté les phrases délexicalisées¹⁴⁶, en précisant dans les instructions qu'il s'agissait d'une langue inconnue des auditeurs afin que ces derniers ne soient influencés d'aucune manière¹⁴⁷. Puis nous avons présenté les vingt-quatre autres phrases dans un ordre volontairement aléatoire (cf. tableau 3 ci-dessous¹⁴⁸).

Nous avons ainsi obtenu des résultats de vingt auditeurs pour les phrases délexicalisées et de dix auditeurs (puisque séparés en deux groupes) pour les autres phrases.

<u>Jastradanz</u>	<u>Gr 1</u>		<u>Gr 2</u>	
P1.B1S01	P3.2S03P	P1.4S04D	P1.9S04P	P2.7S04D
P1.B3S01	P1.4S04	P2.1S02F	P1.8S05P	P3.2S03D
P1.B4S04	P1.8S05D	P1.4S05	P1.1S01P	P1.9S04F
P1.B4S05	P1.3S01	P1.3S01D	P3.1S04D	P1.1S01F
P1.B8S05	P3.1S04F	P3.1S04P	P1.3S01F	P2.7S04
P1.B9S04	P1.9S04D	P2.7S04F	P2.1S02	P1.8S05F
P2.B1S02	P3.2S03F	P1.8S05	P3.1S04	P1.4S05P
P2.B4S01	P2.4S03P	P1.4S05D	P1.4S04P	P2.4S01D
P2.B4S03	P2.4S01F	P2.4S01P	P1.4S05F	P3.2S03
P2.B7S04	P1.9S04	P1.1S01D	P2.4S01	P2.4S03D
P3.B1S04	P2.1S02P	P2.4S03F	P2.1S02D	P1.3S01P
P3.B1S03	P1.1S01	P2.7S04P	P2.4S03	P1.4S04F

Tableau 3 : ordre aléatoire de présentation des stimulus du test perceptif sur signaux manipulés pour les groupes 1 et 2.

¹⁴⁶ Cf. CD ROM joint : *Synthèse voc : Test perceptif : jastra.wav*.

¹⁴⁷ Pour vérifier si les auditeurs soupçonnaient qu'il s'agissait en fait d'anglais, nous leur avons demandé de nous le signaler s'ils reconnaissaient la langue. Un seul auditeur nous a fait part de son impression que c'était du gaélique irlandais...

¹⁴⁸ Cf. CD ROM joint : *Synthèse voc : Test perceptif : groupe1.wav* et *groupe2.wav*.

5.2 Résultats

Comme pour les premiers sondages, présentés dans le chapitre 2, nous avons calculé un pourcentage d'emphasis pour chaque mot à partir du nombre d'auditeurs ayant marqué le mot comme emphatique. Ainsi si 10 auditeurs sur 20 ont souligné un mot, nous disons qu'il est emphatique à 50%. Nous ne leur avons pas demandé de préciser un degré d'emphasis, nous n'obtenons donc qu'un seul nombre pour chaque mot. Pour chaque phrase-source, nous disposons de six séries de pourcentages¹⁴⁹ :

- les pourcentages pour la phrase originale, récupérés des premiers sondages ;
- les pourcentages pour la phrase resynthétisée, mais sans manipulation ;
- les pourcentages de la phrase dont on a modifié la durée des phonèmes ;
- ceux de la phrase dont la F0 a été manipulée ;
- les pourcentages pour la phrase avec les pauses omises ou d'autres ajoutées ;
- enfin ceux de la phrase délexicalisée.

Nous comparerons la phrase originale à la phrase resynthétisée, puis pour toutes les phrases manipulées, c'est la phrase resynthétisée sans modification qui servira de référence.

Grâce à tous ces pourcentages, nous pouvons tirer des conclusions. Nous nous intéresserons à chaque paramètre pris séparément, mais nous allons voir qu'il se révèle extrêmement ardu de les traiter de façon cloisonnée. Nous persisterons tout de même dans ce sens pour débiter, puis ferons une synthèse de toutes nos observations dans la dernière sous-section de ce chapitre.

5.2.1 Phrases originales et phrases synthétiques

Comparons tout de suite les phrases originales et les phrases resynthétisées sans manipulation¹⁵⁰. Les pourcentages ne sont pas exactement les mêmes, bien entendu.

¹⁴⁹ Les séries de pourcentages pour chaque phrase-source sont disponibles sur le CD ROM : *Synthèse voc : test perceptif : résultats*, ou encore en annexe 5d.

¹⁵⁰ Les phrases originales sont données en exemples sonores (exs.58 à 69) sur le CD ROM : *exemples sonores* ; les stimulus synthétiques peuvent être écoutés sur le CD ROM également dans les fichiers

Nous avons déjà précisé que l'on n'arrivait pas à une reproduction parfaite de la phrase originale. D'autre part, nous avons déjà évoqué le manque de contexte ici par rapport aux phrases originales pour lesquelles le contexte était donné grâce à un long morceau que les auditeurs entendaient au préalable.

On remarque qu'en général, l'emphase est moins forte sur les stimulus synthétiques. Seules quelques occurrences font exception, et c'est en général dans les pourcentages moindres d'emphase (autour de 20%). On peut alors penser que les paramètres de F0 et de durée ne suffisent pas à exprimer l'emphase. Dans le segment P1.4S05, « wait » et « see » sont tous deux donnés emphatiques à 77,7% dans la phrase originale et seulement à 10 et 30% respectivement dans la phrase resynthétisée. Le locuteur abaisse sa fréquence fondamentale pour mettre ces deux mots en relief : alors que la moyenne de l'unité intonative se situe à 6,71 demi-tons (147,34 Hz¹⁵¹), « wait » est à -2,86 dt (84,77 Hz) et « see » à 2,84 dt (117,82 Hz). Or, lorsque l'on se trouve dans des fréquences basses, la voix synthétique dans MBROLA est assez mauvaise et sonne artificielle, semblable à celle d'un robot.

De même le « do » du segment P2.7S04, emphatique pour 68,7% des auditeurs dans la phrase originale, tombe à 30% dans la phrase synthétique. Nous pensons que la perception est faussée par la voix synthétique.

Pour « Christmas » (P1.3S01), on passe de 77,7% à 30%. Cela est dû, ici, au fait que l'on ne peut pas intégrer dans MBROLA le paramètre intensité, qui est important en anglais pour la perception de l'accent lexical. Dans le mot « Christmas », par exemple, la première syllabe est plus forte (en intensité sonore) que les autres. On voit nettement sur la figure 39 ci-dessous que le pic d'intensité pour l'UI se situe sur le /I/ de « Christmas », à presque 71 dB :

correspondant aux différentes manipulations (cf. notes 137, 141, 142, 143, 145). Nous redonnons aussi les phrases-sources en annexe 5e afin qu'il soit plus aisé de suivre.

¹⁵¹ Pour la conversion des demi-tons en Hertz, cf. 4.2.3.

Cf. également : <<http://users.utu.fi/jyrtuoma/speech/semitone.html?F1=100&F2=&ST=6%2C71>>

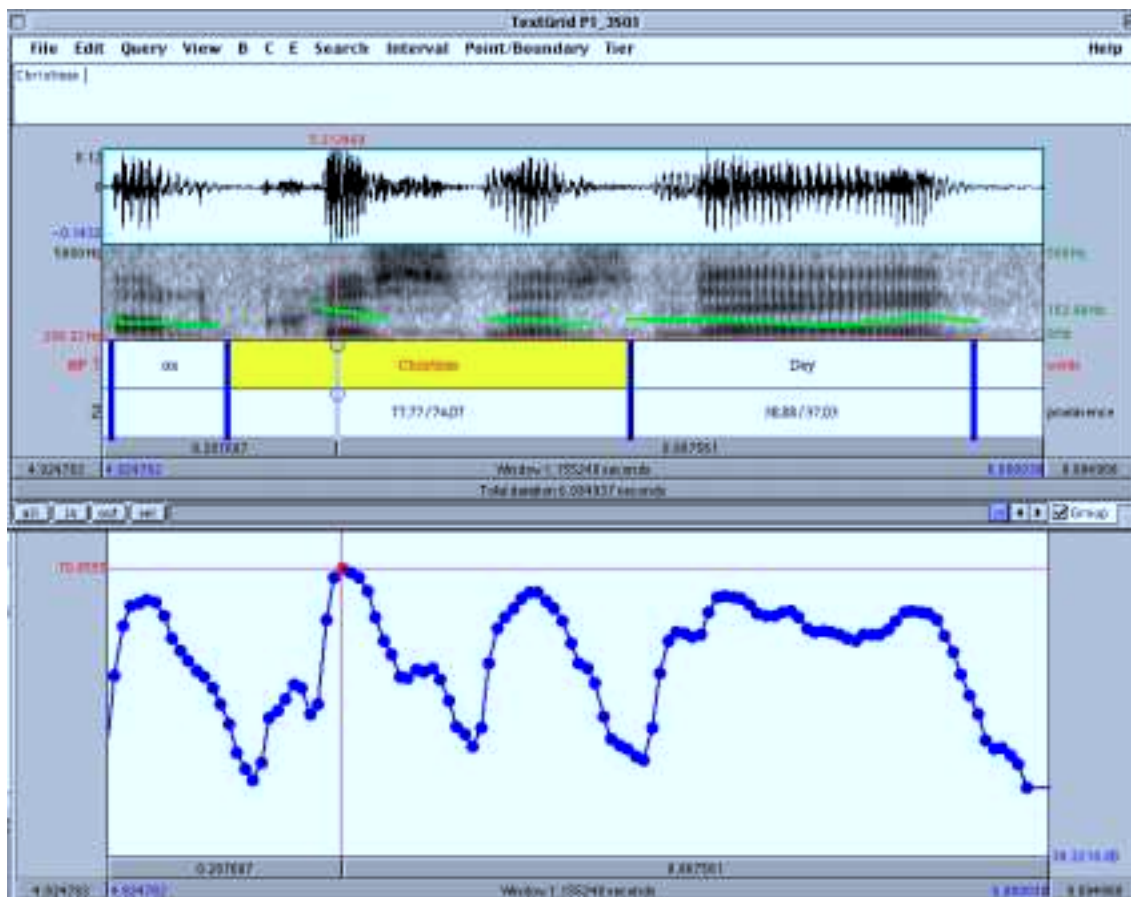


Figure 39 : son, TextGrid et « intensity tier » pour P1.3S01 : « on Christmas Day »

Nous avons également mentionné dans le chapitre 3 quelques cas où nous pensions que l'attaque du mot était capitale dans la perception de l'emphase. Le segment P1.4S04 (☞ **exs.60**) illustre ce point. En effet dans le mot « cabinet », l'attaque du /k/ est particulièrement appuyée, renforcée par un petit coup de glotte, qui rend ce mot saillant par rapport aux autres. De même le mot « government » est prononcé par la locutrice avec un léger coup de glotte. Nous n'avons pu reproduire ces effets dans les phrases resynthétisées, c'est probablement la raison pour laquelle ces deux mots sont perçus moins emphatiques dans les phrases synthétiques : « cabinet » passe de 94,4% à 50%, et « government » de 88,8% à 40%.

Deux autres exemples sont révélateurs de ce phénomène : « ten » (P2.4S03, ☞ **exs.66**) et « just » (P2.7S04, ☞ **exs.67**) sont tous deux moins emphatiques (de 93,7% à 70% pour les deux mots). Or c'est à nouveau l'attaque des phonèmes initiaux qui semble faire la différence.

Cependant, pour la plupart des mots, l'écart n'est pas aussi grand. On obtient deux mots donnés emphatiques par 100% des auditeurs, deux autres par 90% des auditeurs et trois par 80% des auditeurs. Beaucoup de mots sont ensuite donnés par 60 à 70% des auditeurs, ce qui est un bon résultat. Les auditeurs n'ont pas de mal à pointer les emphases.

Quant au contexte, nous avons mentionné plus haut¹⁵² le cas de « domestic violence » (P3.2S03, ¶ exs.69). Nous pensions que « domestic » serait peut-être donné plus emphatique que « violence » du fait du manque de contexte, mais ce n'est pas le cas. Le contexte de la phrase ou de l'unité intonative semble suffire dans la majorité des cas.

5.2.2 Pauses

Nous débutons l'analyse des stimulus manipulés par les pauses car c'est le paramètre qui se révèle le moins pertinent dans la perception de l'emphase. Lorsque nous avons créé les stimulus, nous avons déjà remarqué que l'omission ou l'ajout d'une pause renforçait énormément l'impression d'artificiel. Il est connu qu'un mot est allongé avant une pause, or nous n'avons pas modifié les durées dans les phrases créées pour tester l'importance des pauses. De ce fait, lorsqu'une pause est ajoutée, la durée du mot précédent n'est pas allongée, d'où la sensation d'artificiel renforcée. Lorsqu'une pause est enlevée, le mot précédant n'est pas raccourci non plus, mais cela est moins gênant, on peut allonger un mot sans nécessairement faire une pause après, alors que l'inverse n'est pas vrai. Il s'agit d'une remarque importante selon nous, qui pourrait être prise en compte pour améliorer les futures études de ce type.

En ce qui concerne les résultats de notre test, nous pouvons distinguer deux cas de figure : l'omission d'une pause et l'ajout d'une pause.

5.2.2.1 Omission d'une pause

Nous remarquons que dans la majorité des cas, l'omission d'une pause change très peu la perception de l'emphase du mot qui suit ou qui précède.

¹⁵² Cf. 3.1.2.

Pour quelques cas, l'omission d'une pause rend le mot moins emphatique, comme entre « not » et « even » (P3.1S04) : de 80%, « even » tombe à 30%. Le pourcentage d'emphase chute également pour le mot « last » (P2.1S02) qui perd la pause qui le précédait (de 90% à 60%). Enfin lorsque l'on enlève la pause entre « absolutely » et « freezing » (P2.4S01), le pourcentage d'emphase chute également, de 90 à 60% pour « absolutely » et de 80 à 50% pour « freezing ».

Nous avons même des cas où, paradoxalement, le fait d'enlever la pause rend les mots précédant et suivant plus emphatiques comme pour « a » et « kerfuffle » (P1.1S01). L'article devient emphatique à 20% (contre 0% dans la phrase de référence) et « kerfuffle » à 80% (50% dans la phrase de référence) lorsque la pause entre les deux mots est omise. Nous pensons que cela est dû à l'effet artificiel mentionné ci-dessus. L'allongement de l'article n'est pas choquant avec la pause derrière, mais le devient lorsque la pause est omise. Les auditeurs sont donc interpellés et remarquent davantage les deux mots. Ici un allongement non suivi d'une pause devient un indice d'emphase et ce résultat n'est finalement pas si paradoxal, car le mot allongé est un article, que l'on n'a, en contexte neutre, aucune raison d'allonger, contrairement à un mot comme « absolutely », qui, de par sa nature sémantique, appelle une mise en relief.

Le même phénomène se produit pour « wait » et « see » (dans P1.4S05) : emphatiques pour 10 et 30% des auditeurs respectivement dans la phrase de référence, ils le deviennent tous deux pour 80% des auditeurs lorsque l'on enlève la pause après « see ». Le mot « how » également devient emphatique (40% dans la phrase sans la pause contre 0% dans la phrase non manipulée). Nous pensons que les raisons sont les mêmes que pour le segment P1.1S01, pour « see » et « how » au moins. Quant à « wait », l'argument est moins recevable du fait qu'il n'y a pas de pause après.

Un dernier cas attire notre attention, celui de « domestic violence » (P3.2S03) : nous avons enlevé la pause derrière « violence ». Ce dernier reste emphatique pour 100% des auditeurs, mais « domestic » perd toute emphase. Nous ne voyons aucun argument justifiant ce phénomène.

Dans tous les autres cas, la différence ne nous semble pas significative. Dans le segment P1.3S01, nous avons enlevé la pause qui se trouve après « snow ». Cela ne change rien, comme nous l'avions anticipé, car cette pause se trouve au changement d'UI.

5.2.2.2 Ajout d'une pause

Dans le segment que nous venons de mentionner, nous avons rajouté une pause devant « better », puisque nous pensions que celle qui se trouvait après « snow » n'était pas révélatrice d'une emphase. Le mot « better » est donné à 20% alors qu'il n'est relevé par aucun locuteur dans les autres phrases. Contre toute attente, dans ce même segment, le mot « certainly » est emphatique pour 40% des auditeurs, alors qu'il n'est absolument pas entouré d'une pause, ni dans la phrase originale, ni dans la phrase manipulée, et n'a pas été relevé dans les autres stimulus. Or l'ajout de la pause dans ce segment sonnait artificiel, pour les raisons décrites plus haut. C'est certainement cette impression de « faux » qui a fait qu'un tout autre mot a été donné emphatique.

Dans les autres segments où nous avons rajouté des pauses, les différences ne sont, là non plus, pas très significatives selon nous. Certains mots avant lesquels nous avons rajouté une pause sont même moins emphatiques que dans la phrase non manipulée, à l'exception du « do » du segment P2.7S04. Nous avons enlevé les deux pauses de ce segment et en avons inséré une devant « do », qui passe de 30% dans la phrase de référence, à 50% dans la phrase avec la pause.

5.2.2.3 Conclusion

L'effet des pauses sur l'emphase reste pour nous un phénomène complexe que nous ne parvenons pas à déchiffrer. Nous rejoignons les conclusions tirées en 3.4. Nous pensions que les pauses jouaient un rôle dans la perception de l'emphase et certaines données semblent le confirmer. Mais d'autres données semblent aller à l'encontre de cela. Nous ne disposons probablement pas de données suffisantes. Une étude plus poussée serait nécessaire.

5.2.3 Durées

Nous pouvons tout de suite affirmer que la durée seule ne suffit pas à exprimer l'emphase. En effet, dans toutes les phrases manipulées où nous avons pris pour chaque phonème une durée moyenne, l'emphase est perçue sur certains mots par 100% des

auditeurs. Néanmoins, la durée joue un rôle dans la perception de l'emphase puisque les pourcentages varient entre la phrase non manipulée, qui contient donc les mêmes durées que la phrase originale, et la phrase manipulée.

Nous proposons de comparer les degrés d'emphase attribués aux mots des deux séries de phrases¹⁵³. Nous avons vu dans le chapitre 1 que l'emphase est souvent associée à la durée des phonèmes et qu'un mot emphatique est en général allongé. Nous observerons dans un premier temps si notre étude accrédite cette thèse. Certains travaux antérieurs ont relaté un raccourcissement des mots adjacents à un mot emphatique, et nous verrons si notre étude va dans le même sens. Enfin nous nous intéresserons à un troisième point déjà abordé dans le chapitre 3 (cf. 3.5.2) : l'attaque des mots emphatiques.

5.2.3.1 Niveau du mot

Les mots donnés plus emphatiques dans la phrase manipulée subissent-ils un allongement par rapport à la phrase non manipulée ? Et les mots donnés moins emphatiques sont-ils plus courts ?

Nous avons en effet des cas où cette hypothèse se vérifie, mais on a davantage de cas pour lesquels les mots sont allongés (dans l'original) par rapport à la moyenne (dans la phrase manipulée).

Prenons l'exemple de « snow » (P1.3S01) : il est allongé de 89% par rapport à la moyenne et est beaucoup plus emphatique dans la phrase non manipulée (100% contre 60% dans la phrase manipulée). Tous les phonèmes de ce mot sont allongés, et pas uniquement la voyelle :

	Durée originale (en ms)	Durée moyenne (en ms)
/s/	201	116
/n/	70	65
/@U/	296	120

De la même façon, « are » (P1.9S04) est très long dans la phrase non manipulée (243 ms) et la durée moyenne prise dans la phrase manipulée est de 116 ms. Ce mot est donc

¹⁵³ Cf. annexe 5d ou CD ROM : *Synthèse voc : test perceptif : résultats*. Les durées peuvent être comparées sur les fichiers .pho : *Synthèse voc : Fichiers .pho : Normal* ou *Durée*.

plus emphatique dans la phrase non manipulée : 70% contre 40% dans la phrase manipulée.

Nous pouvons dire la même chose pour « kerfuffle » (P1.1S01), allongé dans la phrase non manipulée de 22% par rapport aux durées moyennes, et donné emphatique par 20% d'auditeurs de plus.

Enfin un dernier exemple est celui de « adults » (P2.7S04), qui est plus emphatique dans la phrase non manipulée que dans l'autre : il est en effet moins long dans la phrase manipulée, et ce pour tous les phonèmes :

	Durée originale (en ms)	Durée moyenne (en ms)
/ɜ/	129	73
/d/	73	52
/V/	87	54
/l/	94	65
/t/	87	66
/s/	167	116

Inversement, le mot « chancellor » (P1.1S01) est plus long dans la phrase manipulée. Ce sont les consonnes en particulier qui sont plus longues : /tS/ passe de 64 ms dans la phrase de référence à 115 ms dans la phrase manipulée, /n/ et /l/ sont aussi plus longs (de 30 et 20 ms respectivement) et le mot est donné emphatique par 50% des auditeurs contre 30% dans la phrase de référence. Il en va de même pour le mot « slow » (P1.9S04) : il est plus long dans la phrase manipulée (10% plus long) et est également pointé par davantage d'auditeurs (100% dans le stimulus manipulé contre 60% dans la phrase non modifiée). Cela pose néanmoins un problème, puisqu'il s'agit de durées moyennes. Pourquoi le mot est-il emphatique ? Bien sûr, la fréquence fondamentale et le contexte ont leur importance, il est difficile d'en faire abstraction. Mais peut-être faut-il aussi regarder les autres mots de la phrase ?

5.2.3.2 Mots adjacents

Les mots qui précèdent « chancellor » sont tous plus courts que dans la phrase non manipulée, à l'exception de « over » (plus long de 12% seulement). Puis l'article « the »

et le mot « chancellor », qui contient une voyelle tendue, sont plus longs que dans la phrase initiale. Ce déséquilibre tend à renforcer l'impression de mise en relief du mot « chancellor ». D'après le contexte, cette mise en avant est inattendue, donc emphatique.

Dans P1.4S05_D, « is » est plus emphatique, et n'est pourtant pas plus long que dans la phrase non modifiée. Il est au contraire bien plus court (115 ms dans la phrase de référence, 63 ms dans la phrase manipulée). Par contre, on remarque que « that » est plus court également (de 116 ms à 73 ms) et se trouve en effet légèrement moins emphatique. Dans la phrase de référence (non manipulée), « that » étant plus long qu'en moyenne, les auditeurs l'ont pointé emphatique, et le « is » suivant est plus long aussi mais de façon assez naturelle, puisqu'il va avec l'allongement de « that ». Dans la phrase manipulée, les moyennes des phonèmes ayant été prises, « that » n'est pas emphatique, et l'on s'attend du fait à entendre le verbe suivant, qui est un auxiliaire, réduit. Comme ce n'est pas le cas, « is » sonne inhabituel et est donc pointé par les auditeurs. Le cas de cette phrase nous paraît intéressant car on perçoit bien ici l'importance des mots qui suivent et précèdent.

L'exemple de P3.2S03 illustre également ce point : « domestic » est plus emphatique dans la phrase manipulée (80%) que dans la phrase de référence (40%), alors que le mot suivant « violence » subit l'effet inverse (100% dans la phrase de départ, 60% dans la phrase manipulée). En effet, « domestic » est allongé par rapport à l'original, c'est surtout la consonne finale qui est plus longue. « Violence » est quant à lui raccourci (en particulier la consonne initiale), bien que suivi d'une pause. On retrouve le problème, mentionné ci-dessus (5.2.2), des pauses qui provoquent un allongement du mot qui les précède. Mais on a également un allongement dans la phrase manipulée de tout ce qui précède « domestic violence », ou plus exactement, tout le début de la phrase est raccourci dans l'original. Or nous avons vu que cela pouvait être également un indice d'emphase. C'est pourquoi « violence » est donné emphatique par 100% des auditeurs dans la phrase de référence et ne l'est plus dans la phrase manipulée. Ici encore, l'importance que prennent tous les mots de la phrase est mise en évidence.

5.2.3.3 Phonèmes

Le segment P2.1S02 propose à nouveau un cas intéressant : « you » est plus emphatique dans la phrase manipulée et il est allongé (de 50 à 70 ms). « Last » est moins emphatique, et est raccourci. Mais la fin de la phrase (« year as well ») est plus emphatique et n'est pourtant pas plus longue si l'on prend les durées globales. Mais on remarque que le /w/ est plus long dans la phrase manipulée (50 ms contre 71 ms). Nous pensons que c'est la raison pour laquelle il est plus emphatique, et nous rejoignons le point soulevé à propos de l'attaque des mots¹⁵⁴, qui est essentielle dans la perception de l'emphase.

Nous avons un cas similaire avec « perfectly » (P1.4S04), qui est nettement moins emphatique dans la phrase manipulée (30%) que dans l'autre (80%). Pourtant, si on prend le mot dans sa globalité, il est plus court dans la phrase non manipulée (de 14%). Mais le phonème initial est, quant à lui, plus long (112 ms pour 85 ms dans la phrase manipulée). C'est la raison pour laquelle « perfectly » est perçu plus emphatique dans la phrase non manipulée. Le début de la phrase (« there's a ») est également raccourci dans la phrase non manipulée (de 50%), ce qui favorise, comme nous l'avons dit plus haut, la perception de l'emphase du mot suivant.

Le cas de « discussed » (P3.1S04) est aussi intéressant pour le niveau du phonème, mais il ne s'agit plus du phonème initial ici. Encore une fois, le mot est moins long dans la phrase manipulée et pourtant plus emphatique. La différence se fait sur un seul phonème, qui n'est pas accentué : le /I/ de la première syllabe. Il est réduit et donc raccourci par le locuteur dans la phrase originale. Il est nettement plus long dans la phrase manipulée (63 ms contre 19 ms dans la phrase normale), et c'est cette disproportion inhabituelle qui le rend saillant.

Nous avons déjà mentionné le mot « domestic » mais nous souhaitons ajouter qu'un phénomène similaire se produit : le phonème final /k/ est plus long dans la phrase manipulée, et certains auditeurs ont souligné la dernière syllabe. Une auditrice nous a également fait remarquer qu'il y avait un déplacement d'accent lexical sur la dernière syllabe de ce mot. Nous avons à nouveau la preuve qu'il suffit de changer la durée d'un phonème pour le rendre plus ou moins emphatique.

¹⁵⁴ Cf. 3.5.2.

5.2.3.4 Conclusion

La durée intervient sans aucun doute dans la perception de l'emphase mais ne représente pour aucun cas de notre étude le seul phénomène pertinent. Si l'on décrète qu'un mot commence à être emphatique lorsqu'il est donné par plus d'un tiers des auditeurs (33%), on n'a aucune occurrence qui passe d'emphatique à non-emphatique uniquement en modifiant les durées. Par contre, associée à d'autres critères que nous allons voir (F0 et sémantisme), la durée contribue à la perception d'une emphase plus ou moins forte.

Cette étude nous a permis de mettre en lumière le fait qu'il ne faut pas se limiter à la durée du mot tout entier, mais que la durée segmentale est capitale. Il suffit d'un allongement sur un seul phonème pour que la perception du mot soit changée. Les mots adjacents ont également leur importance et notre étude va dans le sens des travaux montrant que l'on a tendance à prononcer plus vite les mots qui précèdent un mot que l'on souhaite mettre en relief.

Enfin le point sur lequel nous souhaitons appuyer est celui de l'idée d'inattendu, d'inhabituel. Nous avons employé ces mots tout au long de notre discussion sur les durées et il nous semble que c'est ce qui ressort de cette étude. Lorsqu'une durée est modifiée de façon inattendue, inhabituelle, anormale, le mot qui contient le ou les phonèmes est perçu emphatique.

L'étude de la fréquence fondamentale et des phrases délexicalisées confirmera peut-être cette hypothèse.

5.2.4 Fréquence fondamentale

Contrairement à ce que nous venons de dire à propos de la durée, la fréquence fondamentale semble pouvoir à elle seule faire basculer un mot du non-emphatique à l'emphatique. En effet, dans les phrases manipulées, la F0 est plate, alors que l'on a gardé la F0 originale dans les phrases de référence (resynthétisées non manipulées). Certains mots emphatiques dans les phrases non manipulées ne le sont plus du tout dans les phrases manipulées.

Certains cas sont absolument flagrants.

Le premier cas que nous souhaitons mentionner est celui de « ten-degreet » (P2.4S03). Nous avons vu lors de l'analyse de ce segment que tout le contour intonatif est très plat. Le locuteur emploie un ton très monotone. Or « ten » est pointé par pratiquement tous les auditeurs. Nous nous demandions si une petite montée sur ce mot suffisait à le faire percevoir emphatique. La réponse est positive puisque lorsque la F0 est complètement plate, le mot n'est plus pointé que par 10% des auditeurs (contre 70% dans la phrase non manipulée).

Dans le segment P1.4S05, nous nous attendions à ce que la fréquence fondamentale ait un rôle déterminant dans la perception de l'emphase. La phrase présente en effet un contour intonatif ondulatoire (« sing song ») relativement courant en anglais, suivi d'une grande chute, qui a lieu ici sur le dernier mot de la phrase. La pente sur « currency » est en effet de 79,87 dt/s. Conformément à nos attentes, les mots « wait », « see » (qui correspondent aux extremums du mouvement ondulatoire) et « currency » sont bien moins souvent perçus emphatiques par les auditeurs (de 30 à 0% pour « wait », de 40 à 10% pour « see » et de 60 à 10% pour « currency »).

C'est un contour très courant également que l'on trouve dans P3.2S03, avec une montée préparatoire à une grande chute, qui a lieu ici sur « violence ». Cet exemple est sans équivoque. Alors que pour 100% des auditeurs, « violence » est emphatique dans la phrase non manipulée, seuls 20% d'entre eux perçoivent encore l'emphase lorsque l'intonation est monotone. Une preuve incontestable est également fournie par les résultats sur la phrase délexicalisée : 95% des auditeurs ont jugé emphatique l'équivalent en « jastradanz » de « violence ». Quant à « domestic », il n'est plus du tout pointé emphatique dans la phrase où la F0 est plate (0% contre 40% dans la phrase de référence), c'est donc que la fréquence fondamentale est déterminante en ce qui le concerne aussi. Il n'est pourtant pointé que par 10% des auditeurs dans le stimulus délexicalisé. Nous pensons que le mouvement mélodique est tellement important sur « violence » qu'il en diminue la perception de celui de « domestic ».

De même dans P3.1S04, la fréquence fondamentale est essentielle dans la perception de l'emphase : « even » est perçu emphatique par 80% des auditeurs dans la phrase de référence car il est plus haut que les autres mots de l'UI, avec une moyenne à 12,85 dt (210 Hz) contre 8,82 dt (166,43 Hz) pour l'UI. C'est sur ce mot que l'on trouve le pic maximal de F0 (262 Hz). Il n'est plus pointé que par 10% des auditeurs dans la phrase

manipulée. La durée est plus longue qu'en moyenne sur ce mot (l'allongement est de 65%), et pourtant la perception de l'emphase n'est pas diminuée dans la phrase où les durées sont manipulées. Cela prouve que la fréquence fondamentale seule suffit à exprimer l'emphase dans ce cas.

Dans le segment P1.4S04, nous avons déjà parlé de « perfectly » dont le phonème initial est allongé. C'est également le mot dont le pic maximal de fréquence fondamentale est le plus élevé (221,35 Hz). La moyenne de la F0 sur ce mot est de 8,43 dt (162,73 Hz) alors que la moyenne sur l'UI est de 0,55 dt (103,27 Hz). Cela n'est d'ailleurs pas étonnant puisqu'il s'agit de la tonique de tête de l'unité intonative. Lorsque l'on aplatit la F0, « perfectly », qui était perçu emphatique par 80% des auditeurs dans la phrase non manipulée, n'est plus pointé que par 20% des auditeurs.

Dans le cas de « slow », déjà mentionné plus haut également, nous avons vu que la durée était plutôt moins longue qu'en moyenne dans la phrase non modifiée. Le mot est pointé moins emphatique lorsque la F0 est plate (20%), que dans la phrase de référence (60%), avec la même durée. Ce mot ressort en effet dans le schéma intonatif. C'est le mot le plus haut de l'UI (pic à 245 Hz). Encore une fois, la F0 est révélatrice de l'emphase.

A nouveau, le segment P1.8S05 comprend de larges mouvements mélodiques, sur « in » et « represent » en particulier, qui sont donnés emphatiques par 60 et 70% des auditeurs respectivement dans la phrase non manipulée. Les mouvements mélodiques sont amples mais surtout ils partent de l'aigu, au-delà de 300Hz pour les deux mots, alors que la moyenne de l'UI est à 218 Hz. On observe une chute du pourcentage d'emphase (de 60 à 30% des auditeurs pour « in », et de 70 à 40% pour « represent »). On peut d'ailleurs se demander pour quelle raison « in » est toujours pointé emphatique par certains auditeurs dans la phrase manipulée. Nous remarquons que « in » est en fait beaucoup plus long qu'en moyenne (il subit un allongement de 59% dans la phrase non manipulée, et donc aussi dans la phrase où la fréquence fondamentale est manipulée). Il est étonnant que cette nuance ne soit pas apparue lors de l'analyse du signal où la durée a été modifiée. Avec une durée moyenne, donc bien plus courte, « in » restait emphatique pour 60% des auditeurs (comme dans la phrase non manipulée). Cela nous semble tout à fait révélateur de la combinaison des paramètres dans la perception de l'emphase. Lorsqu'un paramètre disparaît, un autre prend le relais.

Prenons le segment P2.7S04 : les mots « just », « adults » et « do » sont emphatiques dans la phrase non manipulée et ne le sont pratiquement plus dans la phrase manipulée (« just » passe de 70% à 20%, « adults » de 60% à 10% et « do » de 30% à 0%). Or les mesures montrent que les deux premiers mots sont en effet porteurs d'un mouvement mélodique. Pour « just » et « adults », l'écart-type sur le mot est élevé : 4,77 dt (131,7 Hz) et 3,81 dt (124,6 Hz) respectivement. Sur les autres mots du segment, l'écart-type est plutôt autour de 0,5 dt (103 Hz), sauf pour les mots qui sont des toniques de tête d'unités intonatives (comme « think » et « you » ici). On remarque que la pente est assez élevée également : 22,21 dt/s (360,7 Hz/s) pour « just » et 50,81 dt/s (1881 Hz/s) pour « adults ». Le mouvement mélodique sur « do » est beaucoup moins évident (l'écart-type est de 0,34 dt, la pente de 15,44 dt/s et la moyenne du mot est de 12,59 dt pour une moyenne de l'UI à 12 dt). La fréquence fondamentale semble, au vu des pourcentages d'emphase, avoir de l'importance mais nous pensons qu'un lien étroit existe avec le sémantisme, sur lequel nous reviendrons dans la sous-section suivante.

Dans le segment P2.4S01, « absolutely » est donné emphatique par 90% des auditeurs dans le stimulus non manipulé, et par 10% d'entre eux dans la phrase manipulée. La fréquence fondamentale est donc clairement révélatrice. Le mot « absolutely » est en effet porteur d'un mouvement mélodique. L'écart-type en ce qui le concerne est de 5,57 dt, alors qu'il est autour de 0,3 dt pour les autres mots de l'UI, à l'exception de « freezing ». Pourtant « freezing », qui est donc également porteur d'un mouvement mélodique (écart-type = 6,34 dt), reste fortement emphatique dans les deux cas (80% et 90%). Comment expliquer ce phénomène ? N'oublions pas qu'il y a une pause avant « freezing ». Dans la phrase où la fréquence fondamentale a été enlevée, nous avons gardé les pauses d'origine. La pause joue alors certainement le rôle d'amplificateur ici. Nous avons dit à propos des pauses que nous parvenions mal à prouver leur rôle quant à la perception de l'emphase. Il nous semble que ce cas est très intéressant : lorsqu'aucun autre paramètre n'intervient, la pause est déterminante dans la perception de l'emphase. En effet, pour ce segment, le stimulus où l'on a enlevé la pause montre que cette pause-là joue un rôle important. Le sémantisme a aussi certainement une part non négligeable, puisque ces deux mots ont un sémantisme fort. Mais dans la phrase délexicalisée, les deux mots restent emphatiques pour 40% des auditeurs. Le sémantisme n'explique donc pas tout. La durée est peut-être pertinente ici aussi : « freezing » est en effet plus long de

46,59% par rapport à la moyenne. Pour tenter d'expliquer le phénomène de la fréquence fondamentale pour ce segment, nous avons parlé des pauses, de la durée et du sémantisme. Comme nous l'avions annoncé au début de ce chapitre, nous voyons bien qu'il est difficile de traiter chaque paramètre séparément.

Nous avons observé de nombreux cas pour lesquels la fréquence fondamentale semblait essentielle en ce qui concerne l'emphase. Mais ne perdons pas de vue qu'il existe aussi des cas (plus rares, certes) où le fait d'aplatir le F0 n'empêche pas de percevoir une emphase. Quel est alors le rôle que le sémantisme peut jouer ?

5.2.5 Sémantisme

Comme pour les trois critères précédents, on peut affirmer que le sémantisme n'est pas le seul à entrer en jeu dans la perception de l'emphase. La délexicalisation entreprise confirme que sans le sens des mots, on peut percevoir de l'emphase : certaines occurrences ont été données emphatiques à 90 et 95%. Cela prouve de ce fait que les trois autres critères deviennent essentiels.

On note néanmoins que le sémantisme joue un rôle considérable car certaines occurrences jugées fortement emphatiques dans les phrases non manipulées, ne le sont plus lorsque l'on délexicalise.

Nous nous intéresserons dans cette sous-section à la nature des mots. Nous verrons qu'il est à nouveau impossible de parler du sémantisme sans parler des autres critères puisque les phrases délexicalisées gardent la F0 et les durées d'origine. Nous distinguerons deux cas de figure : les mots au sémantisme fort et les mots que nous qualifions de « neutres ».

5.2.5.1 Mots au sémantisme fort

Nous avons choisi certaines de nos phrases-sources parce qu'elles contenaient des mots dont le sémantisme même nous semblait induire une emphase.

C'est le cas dans le segment P2.4S01, qui contient « absolutely freezing ». Ces deux mots, emphatiques pour 90 et 80% des auditeurs respectivement dans la phrase non manipulée, ne le sont plus que pour 40% d'entre eux dans le stimulus délexicalisé. Nous avons vu que dans ce segment la F0 joue un rôle important, mais il semble que la part du sémantisme soit également prépondérante.

Le segment P3.1S04 contient aussi des mots dont le sémantisme est fort : « not » et « even ». C'est l'adverbe qui est le plus emphatique dans la phrase non manipulée (80%). Lorsque l'on délexicalise, le taux d'auditeurs à marquer ce mot emphatique tombe à 20%. Le mot qui correspond à « not » est par contre pointé par 45% des auditeurs, ceci étant probablement dû à la pause qui le suit et donc au fait qu'il est allongé, ou encore à la F0, car « not » est porteur en effet d'un mouvement mélodique aussi. On se rend bien compte ici que lorsqu'un critère disparaît, il est remplacé dans l'oreille de l'auditeur par un autre. Cet exemple est à nouveau une bonne illustration de la combinaison des paramètres car sur la phrase où la F0 est plate, c'est « not » également qui est le plus emphatique (50%). Ceci veut dire que pour rendre « even » emphatique, la nature même du mot ne suffit pas, ni le contexte. Il faut également un mouvement mélodique. S'il manque un de ces deux paramètres, le mot sera perçu nettement moins emphatique, mais un peu tout de même, car un nouveau paramètre viendra prendre le relais, comme la durée, ou la pause.

Le mot « perfectly » est marqué sémantiquement. Il s'agit de l'expression d'un jugement. Nous avons déjà vu que les paramètres F0 et durée étaient essentiels dans la perception de l'emphase de ce mot. Or le critère sémantique ressort aussi : emphatique pour 80% des auditeurs dans la phrase de référence, il ne l'est plus que pour 15%. On aurait ici encore une combinaison des trois paramètres F0, durée (attaque du mot) et sémantisme.

Dans P2.7S04, on trouve un « just » dont le sémantisme est important aussi. Ce n'est pas un mot neutre ici. Là encore, le mot n'est pratiquement plus pointé (10% des auditeurs seulement) dans la phrase délexicalisée.

Tous ces exemples montrent que l'emphase n'est pas obligatoirement toujours exprimée grâce à des moyens acoustiques ou prosodiques. Le sémantisme d'un mot dans une phrase donnée peut suffire. Nous notons néanmoins que les mots au sémantisme fort sont très souvent aussi porteurs d'une mise en relief acoustique ou prosodique.

5.2.5.2 Mots « neutres »

Intéressons-nous aux mots dont le sémantisme n'appelle pas en contexte neutre de mise en relief particulière. Ce sont les mots que nous désignons par le terme « mots neutres ». La phrase P2.7S04 contient un « do » qui n'est normalement pas spécialement marqué. Or il s'agit dans cette phrase du « do » communément appelé le « do » emphatique. C'est-à-dire qu'il n'est pas requis par la syntaxe, et qu'il a été volontairement inséré dans le but de provoquer une insistance, une mise en relief. On a tendance à croire que ce « do » est systématiquement accompagné d'un mouvement mélodique ample. Ce n'est pas le cas ici, comme nous l'avons dit ci-dessus (cf. 5.2.4). Le fait que l'occurrence correspondant à « do » dans le texte transformé en « jastradanz » n'ait été donnée emphatique par aucun auditeur en est la preuve. Nous affirmons que dans un cas comme celui-ci, seul le sémantisme, ou plus exactement la syntaxe, permet de déceler l'emphase. Ceci ne veut pas dire qu'il est impossible d'accentuer mélodiquement le mot, bien au contraire, mais ce n'est pas le cas ici.

Le segment P2.4S03 offre un exemple dans lequel le contexte entier de la phrase est important et pas seulement le mot. Il s'agit du mot « ten », qui n'est pas *a priori* un mot au sémantisme induisant une proéminence particulière. Le contexte est essentiel puisqu'il est question d'une différence de degrés entre Marseille et Londres. Un écart de dix degrés est un gros écart, c'est pour cette raison que « ten » prend de l'importance et est emphatique. Nous avons vu que la fréquence fondamentale est révélatrice dans ce cas. Mais la phrase délexicalisée prouve le contraire puisque l'occurrence correspondant à « ten » n'est donnée emphatique que par 5% des locuteurs. Nous comprenons ici encore l'importance de la combinaison des deux facteurs que sont la mélodie et le contexte.

Un phénomène similaire se produit dans le segment P1.3S01, avec le mot « snow ». Comme « ten », il ne s'agit pas d'un mot dont le sémantisme est particulièrement marqué. Mais dans cette phrase et dans le contexte du discours (le débat politique), « snow » est tout à fait inattendu. On parle de démission du chancelier, et la neige n'a aucun rapport avec cela. C'est bien sûr volontaire de la part du locuteur. Il s'agit d'un procédé oratoire. C'est pourquoi dans un tel contexte, « snow » est emphatique tandis qu'il ne le serait pas forcément dans d'autres situations. On s'attend alors à ne pas le voir pointé dans la phrase délexicalisée et c'est en effet ce qui se produit. Jugé

emphatique par 80% des auditeurs dans la phrase non manipulée, il ne l'est plus que par 15% d'entre eux sans le contexte sémantique.

Dans le segment P1.8S05, le mot « in » est emphatique pour 60% des auditeurs dans la phrase non manipulée. Il est porteur d'un mouvement mélodique, comme nous l'avons vu plus haut (il débute plus haut que les autres mots de la phrase). Mais la fréquence fondamentale seule n'explique pas l'emphase car dans la phrase délexicalisée, le mot n'est plus emphatique (5%). Ce mot n'est effectivement pas un mot sur lequel on anticipe une mise en relief, d'où l'importance du sémantisme. « In » est anormalement souligné par un mouvement mélodique, qui n'a, de ce fait, pas besoin d'être très ample pour être perçu. Mais comme l'auditeur ne s'attend pas à entendre un mouvement mélodique particulier sur ce mot, il le remarque davantage. C'est la raison pour laquelle il n'est pas perçu emphatique dans la phrase délexicalisée.

En P1.9S04, on trouve à nouveau un mot dont la nature ne favorise pas *a priori* une mise en relief : il s'agit de l'auxiliaire « be » (« are » ici). Pour 70% des auditeurs, « are » est emphatique dans la phrase non manipulée. Lorsque l'on aplatit la F0, 60% des auditeurs le donnent encore emphatique, et lorsqu'ils entendent la phrase avec les durées moyennes, 40% le pointent encore. On peut penser que le sémantisme est alors essentiel. Mais 30% des auditeurs donnent le mot qui correspond à « are » emphatique dans la phrase délexicalisée. Voici donc un exemple où tous les paramètres semblent à nouveau se combiner pour rendre « are » emphatique. Dans cette même phrase, le cas de « slow » illustre encore nos propos, de façon un peu moins flagrante cependant. Il s'agit d'un adjectif qui pourrait être dit de façon tout à fait neutre, mais tout le contexte induit l'emphase sur « slow », en opposition avec « are ». La fréquence fondamentale est visiblement plus importante pour cette occurrence, les chiffres le montrent : « slow » est donné emphatique à 60% dans la phrase non manipulée, à 100% dans la phrase dont les durées sont modifiées, à 20% pour la phrase dont la F0 est plate, à 100% quand on rajoute une pause avant « progress », et à 55% dans la phrase délexicalisée.

Dans le segment P2.4S03, comme dans d'autres que nous détaillerons ci-après, ont été pointés lors de l'écoute du stimulus délexicalisé des mots qui n'apparaissaient emphatiques dans aucun autre stimulus, pas même la phrase originale. Nous nous sommes donc intéressée à ce point et avons remarqué qu'il s'agissait souvent de mots

« neutres », et même de mots grammaticaux, comme des prépositions ou des conjonctions. Le cas de ces mots est très intéressant. Dans le segment P2.4S03, toute la fin de la phrase (à partir de « Marseille ») est donnée emphatique par quelques locuteurs (de 20 à 35 % des auditeurs). Ceci n'est pas étonnant pour les équivalents de « Marseille » et « London », mais le devient pour les équivalents de « and and and ». Ils sont donnés emphatiques car on entend un mouvement mélodique en effet et surtout un allongement des phonèmes. Cela est dû à la façon de parler du locuteur, qui hésite ou réfléchit et traîne pour prononcer « London », ce qui le pousse à répéter trois fois « and ». Lorsque l'on a le contexte, on ne peut se méprendre, il ne peut pas s'agir d'une emphase, mais lorsque l'on n'a pas le contexte, la perception est totalement différente. Cela est peut-être aussi dû au fait que la formule de délexicalisation employée ne permet pas de conserver les phénomènes de réduction des voyelles. Toutes les voyelles, y compris /@/, sont en effet remplacées par le même phonème /a/. Dans notre exemple, JC prononce /@n@n@n/ et non /{n{n{n/.

D'autres exemples illustrent le cas des mots « neutres », en particulier le segment P1.3S01. De la même façon, le locuteur réalise sur « at » (en début d'UI) un allongement de 48% par rapport à la moyenne, mais cet allongement n'est pas du tout perçu emphatique dans la phrase non manipulée (le locuteur ne réduit pourtant pas le /{/ en /@/ dans ce segment), à cause du contexte à nouveau. Il n'y a aucune raison d'avoir une emphase sur « at », cela n'aurait aucun sens¹⁵⁵. Sans contexte, l'équivalent de « at » est perçu emphatique par 55% des auditeurs.

Enfin, l'équivalent de « Christmas » dans le stimulus délexicalisé est donné emphatique par 90% des auditeurs alors qu'il ne l'est que pour 30% d'entre eux dans la phrase de référence. La fréquence fondamentale crée ce phénomène car « Christmas » est plus haut (autour de 160 Hz pour le pic maximal) que les autres mots (autour de 120 Hz) et a une pente de 30,6 dt/s. Il est peu emphatique dans la phrase de référence car le mouvement mélodique est naturel, on se trouve au début d'une nouvelle UI et « Christmas » est la tonique de tête. Les changements d'UI sont moins perceptibles lorsque l'on n'a pas le contexte du fait du lien entre la syntaxe et les frontières d'UI.

Ces trois derniers exemples mettent en avant non pas l'importance du sémantisme, mais bien des autres paramètres de fréquence fondamentale et de durée.

¹⁵⁵ Contrairement au « in » ci-dessus, dont la mise en relief est sensée.

Il semblerait alors que le sémantisme soit essentiel pour des mots spéciaux, comme des adverbes d'insistance tels que « even » ou « just », ou bien encore pour des mots qui ne se trouveraient pas dans un contexte neutre, comme le « do » emphatique. Mais les critères de la fréquence fondamentale en particulier, mais également de la durée (dans une moindre mesure) sont couplés au critère du sémantisme, pour faire ressortir des mots que l'on ne s'attend pas à entendre mis en relief. C'est le cas de « slow » ou « are ».

5.3 *Résumé et conclusion*

Après avoir tenté de traiter chaque paramètre isolément, nous proposons un résumé des résultats du test perceptif en reprenant les phrases-sources une par une.

Nous concluons ensuite en faisant une synthèse de toutes nos observations.

5.3.1 Résumé

Au vu des résultats du test¹⁵⁶, nous pouvons extraire le paramètre qui semble le plus important dans la perception du ou des mot(s) le(s) plus emphatique(s) des phrases de référence (resynthétisées non manipulées) de chaque segment.

- P1.1S01 : « kerfuffle » et « resign » ne sont tous deux donnés emphatiques que par 50% des auditeurs dans la phrase non manipulée. Aucun critère ne se dégage vraiment. Contrairement à nos attentes (pour « kerfuffle » en particulier), les deux mots sont même emphatiques pour 65% des auditeurs dans le stimulus délexicalisé, ce qui tend à prouver que ce sont plutôt les critères de durée, de F0 et la pause qui sont pertinents dans ce segment. L'allongement de l'article « a » et la pause devant « kerfuffle » sont déterminants comme nous l'avons vu. L'équivalent de « a » en « jastradanz » est en effet pointé par 25% des auditeurs (ceci étant peut-être aussi dû au fait qu'on ne peut percevoir la réduction de la voyelle dans la phrase délexicalisée).

- P1.3S01 : l'emphase sur « snow » provient principalement du contexte et de façon moindre de la durée (allongement).

- P1.4S04 : pour « perfectly », F0, durée et sémantisme sont combinés, avec une légère prépondérance pour le sémantisme ; si on les dissocie, le mot devient nettement moins emphatique. Les autres mots de cette phrase ne sont que peu emphatiques dans le stimulus resynthétisé non manipulé.

- P1.4S05 : dans ce segment la fréquence fondamentale est très nettement le critère le plus important pour l'emphase. Pour « see », la durée est ajoutée.

¹⁵⁶ Cf. annexes 5d pour les résultats et 5e pour les phrases-sources.

- P1.8S05 : on observe une combinaison des paramètres fréquence fondamentale et sémantisme pour « in » et c'est la fréquence fondamentale qui l'emporte pour « represent ».
- P1.9S04 : la F0 est prépondérante pour « slow » ; c'est plutôt le contexte, combiné à un allongement du phonème, qui joue dans l'emphase sur « are ».
- P2.1S02 : nous n'avons pas évoqué le cas de « last », emphatique pour 90% des auditeurs dans la phrase de référence. La fréquence fondamentale fait baisser le pourcentage à 50%, comme la délexicalisation. Dans la phrase avec les durées moyennes, on a 70% et lorsque l'on enlève la pause, 60%. Aucun paramètre ne semble vraiment flagrant. Nous pensons à une combinaison de tous les paramètres, l'un prenant le relais de l'autre, comme nous l'avons déjà expliqué.
- P2.4S01 : ici encore une combinaison de la durée (pour « freezing » principalement), de la F0 et du sémantisme rendent les deux mots « absolutely » et « freezing » emphatiques à un haut degré.
- P2.4S03 : le contexte est essentiel dans ce segment, et si la fréquence fondamentale lui est associée, le mot est très emphatique. Si l'on ne garde que la mélodie, alors on perd toute emphase. Si l'on ne garde que le contexte, le mot est aussi nettement moins emphatique. Ici les deux paramètres sont indissociables pour créer l'emphase, contrairement au phénomène de relais mentionné pour « last » par exemple.
- P2.7S04 : fréquence fondamentale et sémantisme sont à nouveau associés pour « just » et la pause le précédant semble prendre de l'importance aussi. Pour « adults », la fréquence fondamentale est prépondérante et pour « do » c'est la syntaxe.
- P3.1S04 : l'emphase sur « even » est rendue grâce à une combinaison de la mélodie et du contexte.
- P3.2S03 : la F0 est cruciale pour la perception de l'emphase de ce segment.

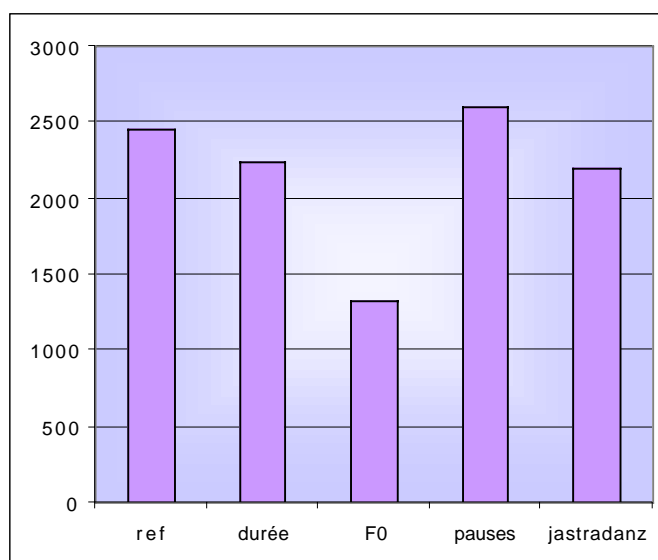
5.3.2 Conclusion

Nos hypothèses de départ quant à l'importance de la mélodie, de la durée des phonèmes et du contexte sémantique sont vérifiées par l'étude décrite dans ce chapitre. En ce qui concerne les pauses, la question reste ouverte.

L'interprétation des résultats montre que l'on peut difficilement parler d'un paramètre sans en aborder un autre. Cela prouve combien tous les paramètres sont imbriqués les uns dans les autres.

Cependant, parmi les quatre critères testés ici, la fréquence fondamentale est le critère qui ressort le plus souvent dans la perception de l'emphase.

Pour chaque série de stimulus, nous avons additionné les pourcentages d'emphase de chaque mot. Cela nous donne le graphique suivant :



Graphique 11 : pourcentage cumulé d'emphase pour toutes les séries de stimulus

Ce graphique montre très nettement que c'est dans les stimulus où la fréquence fondamentale est modifiée que les pourcentages chutent.

Le critère sémantique vient en seconde position, puis c'est la durée.

La délexicalisation a permis de prouver l'importance du contexte dans la perception de l'emphase, mais également d'affirmer qu'il n'est que de très rares cas où le sémantisme (ou la syntaxe) est seul à exprimer l'emphase.

Nous pouvons également affirmer au terme de cette étude sur signaux manipulés que l'emphase est un phénomène complexe. On ne peut extraire un seul paramètre acoustique ou prosodique en corrélation avec l'emphase.

Nous avons vu dans de rares cas qu'un seul paramètre suffisait à rendre un mot emphatique ; dans d'autres cas, un paramètre est prépondérant mais d'autres sont sous-

jacents et ont leur importance quant au degré d'emphase ; enfin dans un troisième type d'occurrences, deux paramètres (il s'agit toujours de la F0 et du sémantisme) sont nécessaires et indissociables dans la perception de l'emphase.

L'emphase semble donc être perçue grâce à une combinaison subtile et surtout variable de plusieurs paramètres.

Conclusion

Cette étude avait pour objectif de déterminer quels sont les paramètres acoustiques et prosodiques les plus révélateurs de l'emphase dans le discours spontané anglais. Nous avons choisi de réaliser une étude principalement expérimentale.

Mais qu'est-ce que l'emphase ? Nous sommes volontairement partie d'une définition vague de ce concept, mettant en relief les notions d'implication plus importante de la part du locuteur que lors d'un discours neutre, ainsi que les notions d'inhabituel ou d'inattendu.

Base de données

Nous avons choisi un corpus de parole spontanée, regroupant trois types de discours : du débat politique, de la conversation informelle, et une émission de radio où deux femmes traitent d'un sujet émotionnel qui les touche particulièrement. L'étude de la parole naturelle s'inscrit en effet dans la tendance actuelle, la parole de laboratoire ayant déjà été largement explorée. Elle a de plus ses limites : les stimulus étant délibérément choisis pour les besoins d'une étude particulière, les résultats de cette dernière risquent de s'en trouver quelque peu faussés. Lorsque l'on étudie le discours spontané, on se trouve néanmoins face à des difficultés telles que la qualité de l'enregistrement ou les chevauchements, pour ne donner que quelques exemples. Pour y pallier, nous avons sélectionné certains passages qui constituent notre corpus primaire, sans prendre les morceaux où plusieurs personnes parlent en même temps par exemple.

Méthode

Afin de déterminer quels étaient les passages ou mots emphatiques de notre corpus primaire, nous avons opté pour une étude perceptive : nous avons réalisé des sondages auprès d'auditeurs anglophones naïfs, leur demandant de pointer les passages ou mots

qui leur semblaient emphatiques dans les morceaux écoutés. Grâce à ces sondages, nous avons pu déterminer un degré d'emphase pour chaque mot, fondé sur le pourcentage d'auditeurs ayant pointé le mot. Cet « indice d'emphase » a servi de point de départ à nos analyses acoustiques. Cette méthode nous semblait la mieux adaptée, étant la plus objective. Nous avons en effet voulu nous baser sur ce qui est intuitivement perçu par des auditeurs naïfs et non imposer une idée préconçue de ce qu'est l'emphase en évitant ainsi une démarche circulaire.

Préparation du corpus

Ayant rassemblé nous-même une base de données, nous en avons effectué manuellement l'étiquetage en mots, puis en unités tonales¹⁵⁷ (unité rythmique que nous avons largement commentée), puis en unités intonatives. Tout ce travail a été réalisé grâce au logiciel PRAAT (Boersma & Weenik, 1996). Nous avons également transcrit, automatiquement cette fois avec seulement quelques corrections manuelles, tous les segments de notre corpus primaire en phonétique.

Analyse acoustique

Nous avons entrepris ensuite de rechercher les paramètres les plus pertinents dans l'emphase. A cette fin, nous avons tout d'abord effectué une analyse acoustique de nos segments, en émettant l'hypothèse que quatre critères étaient certainement essentiels dans la perception de l'emphase : la fréquence fondamentale, la durée segmentale, l'intensité globale et les pauses. Nous avons choisi ces quatre critères au vu des études antérieures.

Cette analyse acoustique nous a permis de vérifier dans un premier temps nos hypothèses et de dégager ensuite des paramètres qui nous paraissaient importants, dans le but de mettre en œuvre une analyse statistique.

Analyse statistique

Nous avons choisi d'effectuer une analyse objective grâce au logiciel de statistiques CRUISE (Kim & Loh, sous presse), un outil récent et très performant. Ce logiciel de statistiques de type CART génère des arbres de classification qui permettent de

¹⁵⁷ Que nous définissons comme suit : un mot accentué suivi de tous les mots inaccentués jusqu'au suivant (cf. 3.1.2.2).

déterminer l'importance relative de chaque paramètre. Nous avons réalisé l'analyse statistique sur les segments du corpus et avons aussi fait une analyse par locuteur et par type de discours, afin de voir si des stratégies se dégageaient.

Etude sur signaux manipulés

Enfin, une étude sur signaux manipulés grâce au synthétiseur de parole MBROLA (Dutoit *et al.*, 1996 ; Dutoit, 1997) a constitué le pendant de l'analyse statistique. Nous avons modifié à tour de rôle la fréquence fondamentale, les durées segmentales et les pauses dans des segments de notre corpus primaire. Nous avons également délexicalisé ces segments et mis en place un nouveau test de perception auprès d'anglophones naïfs sur le même principe que le test de départ. Ce test perceptif sur signaux manipulés nous a permis d'intégrer le critère du sémantisme que nous n'avions pu insérer dans nos paramètres pour l'analyse statistique. Nous n'avons, en revanche, pas pu tester l'intensité.

Résultats et discussion

Nous avons décrit dans le chapitre 4 tout le travail préliminaire à l'analyse statistique finale. En effet, afin d'optimiser l'utilisation du logiciel CRUISE, nous nous sommes livrée à de nombreux essais. Nous avons vu qu'il semblait essentiel de déterminer des coûts de classification erronée, grâce auxquels nous avons pu pencher, à l'instar de Ladd *et al.* (1994), pour une notion dichotomique de l'emphase, avec des degrés dans l'emphatique.

Cependant, avant de nous lancer dans les coûts de classification erronée, nous avons déterminé des coefficients d'emphase et avons réalisé différents essais avec des coûts de classification erronée égaux. C'est à partir de cela que nous avons opté pour un coefficient d'emphase avec seulement trois classes. Nous n'avons pas essayé les coefficients comptant davantage de classes avec des coûts de classification erronée inégaux. Nous pensons qu'il serait peut-être intéressant de procéder à une telle analyse. En ce qui concerne les paramètres que nous avons sélectionnés pour l'analyse statistique, nous regrettons de n'avoir pas été en mesure d'en intégrer certains.

Nous pensons principalement à l'analyse de la balance spectrale. Il serait intéressant de voir si ce paramètre ressort parmi les premiers dans les arbres de classification. En effet,

les chercheurs ayant analysé ce phénomène (cf. Sluitjer & van Heuven, 1993 ; 1996 ; Campbell, 1995 ; Campbell & Beckman, 1995) en arrivent à la conclusion que les fréquences hautes du spectre augmentent lorsqu'il y a emphase.

Un autre critère essentiel est celui du type de mouvement mélodique qui précède le mot emphatique. Nous avons mentionné ce phénomène, mais n'avons pas pu intégrer un paramètre s'y rapportant. Nous avons tenté de déterminer des paramètres qui rendent compte de la mélodie générale de l'unité intonative tout entière (pente sur l'UI : *MAS_UI* ; écart-type sur l'UI : *SD_UI*). Ces paramètres ne sont malheureusement pas assez précis. On ne peut pas savoir si l'on a un contour montant ou descendant par exemple. De plus, nous pensons que ce qui précède uniquement le mot pointé emphatique a peut-être de l'importance. Nous avons manqué de temps pour mesurer ce critère car il était impossible de le faire automatiquement pour chaque mot du corpus. Il aurait fallu faire la plupart des analyses manuellement sur près de 2400 occurrences. Nous pensons cependant qu'il y a là une possibilité d'améliorer notre étude.

L'analyse des pauses mérite aussi certainement d'être complétée. Il nous a été très difficile de tirer des conclusions sur le rôle des pauses. Nous avons parlé des pauses entourant les mots emphatiques et avons vu qu'elles pouvaient ne pas se trouver juste avant ou juste après l'emphase, mais décalées d'un mot ou deux. Cela n'a pas été pris en compte dans l'analyse statistique. De même nous n'avons pas intégré le paramètre de la durée des pauses pour l'analyse statistique. Enfin, les pauses sont de toute évidence liées à la durée du mot qui les précède, et nous n'avons pas pris ce phénomène en compte, en particulier pour l'étude sur signaux manipulés.

En ce qui concerne cette dernière étude, nous déplorons un trop petit nombre d'auditeurs. Les psycho-linguistes estiment que l'on peut raisonnablement fonder l'étude d'une variable sur dix locuteurs. Nous pensons que c'est trop peu. Ayant divisé notre test de perception en deux pour le confort des auditeurs, nous n'avons eu que dix auditeurs pour chaque phrase manipulée (à l'exception des phrases délexicalisées pour lesquelles nous en avons eu vingt). Nous aurions peut-être obtenu des résultats légèrement différents avec davantage d'auditeurs.

Malgré toutes ces remarques induisant des possibilités d'amélioration de notre étude et de recherches futures, nous sommes parvenue à des résultats satisfaisants. L'analyse

statistique et l'étude sur signaux manipulés nous ont permis de dégager des critères récurrents dans l'emphase.

L'énergie dépensée pour prononcer un mot (que nous avons appelée, à la suite de Beckman (1986), « amplitude totale ») est un paramètre fondamental dans la perception de l'emphase en discours spontané. L'énergie est en fait une combinaison de la durée et de l'intensité (cf. 4.2.6). Ainsi, si l'on veut mettre un mot en relief de façon emphatique, on peut le prononcer avec une intensité plus forte que sur les autres mots de l'énoncé, ou encore l'allonger. On peut bien sûr combiner les deux aspects.

La fréquence fondamentale est un autre corrélat de l'emphase, qui peut être associé à l'énergie ou la remplacer pour exprimer l'emphase. C'est le cas lorsque l'on se trouve face à des mots courts, monosyllabiques par exemple, dont l'amplitude totale sera nécessairement faible, la durée d'une syllabe, même allongée, ne pouvant pas atteindre celle de plusieurs syllabes. La fréquence fondamentale prend alors le relais. Elle se manifeste par la hauteur mélodique, que l'on augmente en général sur le mot emphatique, ainsi que par un mouvement mélodique souvent ample sur le mot emphatique.

Enfin, le contexte et le sémantisme se sont également révélés cruciaux dans la perception de l'emphase, comme nous l'a montré la délexicalisation. Nous disions en introduction à ce travail qu'au-delà des mots, la prosodie joue un rôle fondamental dans la parole et dans la perception de l'emphase. Notre étude confirme ce propos mais nous pouvons désormais affirmer que les mots aussi sont essentiels dans la perception de l'emphase. Encore une fois, il nous semble judicieux d'employer les mots « relais » et « combinaison ». Une combinaison de plusieurs facteurs entre en jeu dans l'emphase, ou bien un critère déterminant prend le relais d'un autre selon les cas. Les critères acoustiques et prosodiques suffisent parfois à faire percevoir une emphase – un seul critère est parfois suffisant, plusieurs paramètres combinés peuvent se révéler nécessaires. Dans d'autres cas, le sémantisme est essentiel et la perception de l'emphase peut même être faussée si l'on perd le contexte sémantique. Un mot peut sonner emphatique uniquement grâce à un procédé syntaxique.

Au terme de ce travail, nous pouvons donc affirmer que l'emphase dans le discours spontané anglais se manifeste au niveau acoustique par une énergie plus importante que sur les mots adjacents, et au niveau prosodique par un mouvement mélodique qui part

de plus aigu et est plus ample que sur le reste de l'énoncé. Lorsque l'on applique ces critères sur des mots qui ne se prêtent pas *a priori*, au niveau contextuel ou sémantique, à une mise en relief, l'emphase n'en est que plus importante. C'est à ce niveau que le sémantisme et le contexte jouent un rôle crucial.

Enfin, nous avons vu que l'emphase est en réalité bien souvent exprimée grâce à une combinaison de tous ces paramètres.

On pourrait nous reprocher de fonder partiellement nos conclusions sur une analyse statistique qui ne donne finalement pas de très bons résultats, les taux de prédiction pour les classes emphatiques étant assez faibles. C'est la raison pour laquelle nous avons parlé tout au long de notre étude de stratégies personnelles, propres à chaque locuteur, les cas particuliers décrits en 3.5 (tels que l'attaque des mots et les coups de glotte, ou une baisse de l'intensité, par exemple) en étant l'exemple indéniable. Cependant, même s'il est vrai que l'emphase peut être exprimée de façons diverses, nous pensons que les critères mentionnés ci-dessus représentent des constantes fondamentales.

Voici donc notre définition de l'emphase :

L'emphase est une mise en relief inattendue qui marque une implication de la part du locuteur plus importante qu'à la normale. Cette mise en relief est obtenue en général grâce à une augmentation de l'énergie dépensée pour prononcer l'occurrence emphatique, et/ou une augmentation de la hauteur et/ou du mouvement mélodique sur cette occurrence.

Apport de l'étude et perspectives de recherche future

Notre travail s'inscrit dans un cadre de recherche dont les applications pourront peut-être se révéler utiles pour la rééducation de certains malades ou malentendants. Nous pensons plus particulièrement à certaines pathologies comme des désordres dysprosodiques consécutifs à une atteinte cérébrale (cf. Louis, en cours), dans lesquelles les malades parlent de façon agressive, par exemple, sans s'en rendre compte. Les désordres peuvent être liés à la perception ou à la production, ou parfois les deux. Le fait de visualiser des courbes d'intensité ou de fréquence fondamentale dans lesquelles on a appliqué les critères de l'emphase, ainsi que la courbe du malade, pourra peut-être aider à mieux comprendre les difficultés du patient, et à améliorer sa rééducation.

Nous pensons également que les recherches effectuées pourront contribuer à améliorer la parole synthétique et la reconnaissance automatique, dans le cadre du projet ProZed par exemple (Hirst, 2000). Ce projet vise à développer des règles de prosodie pour une amélioration de la parole de synthèse. Il se fonde sur quatre niveaux de représentation, suivant Hirst, Di Cristo & Espesser (2000), à savoir :

- un niveau physique,
- un niveau phonétique,
- un niveau phonologique de surface,
- un niveau phonologique sous-jacent.

Le niveau phonologique sous-jacent est l'interface entre la forme phonologique et l'interprétation sémantique/syntaxique.

ProZed s'intéresse plus particulièrement aux deux niveaux intermédiaires en fournissant des outils d'étiquetage automatique afin de pouvoir tracer des chemins à deux sens entre les données acoustiques et les représentations phonétiques d'une part et entre les représentations phonétiques et les représentations phonologiques d'autre part.

Ce programme de recherche ne vise donc pas dans un premier temps à prédire les caractéristiques prosodiques d'un énoncé, mais plutôt à reproduire ces caractéristiques de la façon la plus naturelle possible.

C'est dans ce cadre que nous espérons intervenir grâce à notre travail en fournissant des paramètres propres à l'emphase.

Enfin, nous pensons que notre étude pourra peut-être aussi trouver des applications dans l'enseignement de l'anglais. Il est plus aisé d'expliquer à un apprenant comment placer une emphase grâce à des critères acoustiques, prosodiques et sémantiques. Il serait intéressant de faire une étude comparée de l'emphase en français, afin de voir si les mêmes critères ressortent. La délexicalisation entreprise nous a laissé entrevoir l'importance du contexte, du sémantisme et de la syntaxe. Nous aimerions approfondir cet aspect en particulier en comparaison avec le français. Une analyse syntaxique des deux langues permettrait peut-être de comprendre encore mieux le phénomène de l'emphase.

Références bibliographiques

- Abercrombie, D., 1964, Syllable quantity and enclitics in English, in D. Abercrombie *et al.*, (eds.) : 216-222.
- Abercrombie, D., 1976, « Stress » and some other terms, *Work in Progress* 9, Edinburgh University, Department of Linguistics, 51-53.
- Abercrombie, D., Fry, D.B., McCarthy, P.A.D., Scott, N.C., Trim, G.L.M., (eds.), 1964, *In Honour of Daniel Jones*, London : Longman.
- Armstrong, L. & Ward, I., 1926, *A Handbook of English Intonation*, Cambridge : Heffer & sons.
- Arons, B., 1994, Pitch-based emphasis detection for segmenting speech recordings, *Proceedings of the International Conference of Spoken Language Processing* (September 18-22, Yokohama, Japan), 4 : 1931-1934.
- Arrivé, M., Gadet, F., Galmiche, M., 1995, *Grammaire d'aujourd'hui, Guide alphabétique de linguistique française*, Paris : Flammarion.
- Asher, R.E., (ed.), 1994, *The Encyclopaedia of Language and Linguistics*, Oxford : Pergamon Press.
- Astésano, C., Espesser, R., Hirst, D.J., Llisterri, J., 1997, Stylisation automatique de la fréquence fondamentale : une évaluation multilingue, *Actes du 4^{ème} Congrès Français d'Acoustique*, Marseille (14-18 avril 1997) : 441-443.
- Auran Cyril, en cours, *Prosodie et interprétation pragmatique*, Thèse de Doctorat, Laboratoire Parole et Langage, Aix-en-Provence : Université de Provence.
- Ayers, G.M., 1995, Nuclear pitch accent types and prominence : some psycholinguistic experiments, *Proceedings of the 13th International Congress of Phonetic Sciences*, Stockholm, Suède, vol 3 : 660-663.
- Bagou, O., 1998, *L'implication emphatique : étude de la narration orale spontanée*, Mémoire de DEA, Laboratoire Parole et Langage, Aix-en-Provence : Université de Provence.
- Bagshaw, P.C., 1993, An investigation of acoustic events related to sentential stress and accents in English, *Speech Communication*, 13 : 333-342.
- Barnwell, T.P., 1971, An algorithm for segmental durations in a reading machine context, *Technical Report 479*, MIT, Research laboratory of Electronics.

- Barry, W.J., 1995, Phonetics and phonology of speaking styles, *Proceedings of the XIIIth International Congress of Phonetic Sciences* (Stockholm 1995), 2 : 4-10.
- Beckman, M., 1986, *Stress and Non-Stress Accent*, Dordrecht : Foris Publications.
- Beckman, M., 1997, A typology of spontaneous speech, in Y. Sagisaka *et al.*, (eds.) : 7-26.
- Beckman, M. & Edwards, J., 1994, Articulatory evidence for differentiating stress categories, in Keating, P.A., (eds.), *Phonological Structure and Phonetic Form. Papers in Laboratory Phonology III*, Cambridge : Cambridge University Press : 1-33.
- Bloomfield, L., 1933, *Language*, New York, Holt, London : Allen & Unwin, 1935.
- Boersma, P. & Weenik, D., 1996, PRAAT : a system for doing phonetics by computer, *Report of the Institute of Phonetic Sciences of the University of Amsterdam*, 132.
- Bolinger, D., 1958, A theory of pitch accent in English, *Word* 14, 109-149, reprinted in Abe, I. & Kanekiyo, T., (eds.), 1965, *Forms of English : Accent, Morpheme, Order*, Harvard : Harvard University Press.
- Bolinger, D., 1961, « Contrastive accent and contrastive stress », *Language*, 37 : 83-96.
- Bolinger, D., 1972, « Accent is predictable (if you're a mind reader) », *Language*, 48 : 633-344.
- Bolinger, D., 1986, *Intonation and its Parts*, Stanford, CA : Stanford University Press.
- Bolinger, D., 1989, *Intonation and its Uses*, London : Edward Arnold.
- Botinis, A. (ed.), 2000, *Intonation. Research and Applications*, Dordrecht : Kluwer Academic Publishers.
- Bouscaren, J., Chuquet, J. & Danon-Boileau, L., 1992, *Introduction to a Linguistic Grammar of English ; an Uttered-centered approach*, Gap, Paris : Ophrys.
- Bouzon, C. & Blanchard M., 2000, *The influence of prosodic factors on the duration of phonemes*, Mémoire de Maitrise, Laboratoire Parole et Langage, Aix-en-Provence : Université de Provence.
- Bouzon, C., 2001, *L'influence des facteurs prosodiques sur la durée des mots en anglais britannique*, Mémoire de DEA, Laboratoire Parole et Langage, Aix-en-Provence : Université de Provence.
- Bright, W. (ed.), 1992, *International Encyclopedia of Linguistics*, New York, Oxford : Oxford University Press.
- Brown, G., Currie, K.L. & Kenworthy, J., 1980, *Questions of Intonation*, London : Croom Helm.
- Brown, W.S. & McGlone, R.E., 1974, Aerodynamic and acoustic study of stress in sentence productions, *Journal of the Acoustical Society of America* 56 : 971-974.
- Bussman, H., Trauth, G.P., Kazzazi, K., 1996, *Routledge Dictionary of Language and Linguistics*, London and New York : Routledge.

- Caffi, C. & Janney, R.W., 1994, Towards a pragmatics of emotive communication, *Journal of Pragmatics*, 22 : 325-373.
- Campbell, N., 1992a, Syllable-based segmental duration, in Bailly, G. & Benoît, C., (eds.), *Talking Machines*, Amsterdam : North-Holland, 211-224.
- Campbell, N., 1992b, *Multi-level Timing in Speech*, Ph.D. Thesis, Sussex University.
- Campbell, N., 1992c, Prosodic encoding of English speech, *Proceedings ICSLP'92*, Banff, Vol 1 : 663-666.
- Campbell, N., 1995a, Loudness, spectral tilt, and perceived prominence in dialogues, *Proceedings of the XIIIth International Congress of Phonetic Sciences* (Stockholm 1995), 3 : 676-679.
- Campbell, N., 1995b, Prosodic influence on segmental quality, *Proceedings Eurospeech '95*, Madrid, Vol 2 : 1011-1014.
- Campbell, N., 1997, Synthesizing Spontaneous Speech, in Y. Sagisaka *et al.*, (eds.) : 165-186.
- Campbell, N., 2000, Timing in speech : a multi-level process, in M. Horne, (ed.) : 281-334.
- Campbell, N. & Isard, S.D., 1991, Segment duration in a syllable frame, *Journal of Phonetics* 19 : 37-47.
- Campbell, N. & Beckman, M., 1995, Stress, Loudness and Spectral Tilt, *Proceedings of the Acoustical Society of Japan*, Spring meeting, 3-4-3.
- Cavé C., Guaitella I., Bertrand R., Santi S., Harlay F., Espesser R., 1996, About the relationship between eyebrow movements and F0 variations, *Proceedings of the ICSLP*, Philadelphie, 2175-2178.
- Chafe, W., 1970, *Meaning and the Structure of Language*, Chicago : Chicago University Press.
- Chafe, W., 1976, Givenness, contrastiveness, definiteness, subjects, topics, and point of view, in Li, C. N., *Subject and Topic*, New York : Academic Press, 25-55.
- Chafe, W., Dubois, J., Thompson, S., 1991, Towards a New Corpus of Spoken American English, in Aijmer, K. & Altenberg, B., (eds), *English Corpus Linguistics*, London : Longman.
- Chomsky, N. & Halle, M., 1968, *The Sound Pattern of English*, New York, Evanson & London : Harper and Row.
- Chomsky, N., 1969, *Deep Structure, Surface Structure and Semantic Interpretation*, IUCL reprinted in Chomsky, N., 1972, *Studies on Semantics in Generative Grammar*, The Hague : Mouton.
- Classe, A., 1939, *The Rythm of English Prose*, Oxford : Blackwell.

- Clear, J., 1993, The British National Corpus, in Landow, G.P. & Delaney, P. (eds.), *The Digital Word : Text-based computing in the humanities*, Cambridge, Mass : MIT Press, 163-167.
- Coker, C., Umeda, N., Browman, C., 1973, Automatic synthesis for ordinary English text, *IEEE Trans. Audio Electroacoust.* AU-21 : 293-297.
- Coleman, H.O., 1914, Intonation and Emphasis. *Miscellanea Phonetica* 1 : 6-26.
- Collinge, N.E., (ed.), 1990, *An Encyclopaedia of Language*, London and New York : Routledge.
- Cooper, W. & Sorensen, J., 1981, *Fundamental Frequency in Sentence Production*, New-York, Heidelberg, Berlin : Springer-Verlag.
- Cooper, W.E., Eady, S.J., Mueller, P.R., 1985, Acoustical aspects of contrastive stress in question-answer contexts, *Journal of the Acoustical Society of America* 77/6 : 2142-2156.
- Coulthard, M. & Brazil, D., 1982, The place of intonation in the description of interaction, in Tannen, D., (ed.), *Analizing Discourse : Text and Talk*, 94-112 (GURT 1982), Washington, DC : Georgetown University Press.
- Couper-Kuhlen, E., 1986, *An Introduction to English Prosody*, Tübingen : Niemeyer.
- Couper-Kuhlen, E. & Selting, M., (eds.), 1996, *Prosody in Conversation. Interactional Studies*, Cambridge : Capmbridge University Press.
- Croft, W., 1995, Intonation units and grammatical structure, *Linguistics*, 33/55 : 839-882.
- CRUISE, disponible sur : <<http://www.wpi.edu/~hkim/cruise/>>
- Cruttenden, A., 1986, *Intonation*, Cambridge : Cambridge University Press.
- Crystal, D., 1969, *Prosodic Systems and Intonation in English*, Cambridge : Cambridge University Press.
- Crystal, D., 1975, *The English Tone of Voice : Essays on Intonation, Prosody and Paralanguage*, London : Edward Arnold.
- Crystal, D., 1993, *The Cambridge Encyclopedia of Language*, Cambridge : Cambrige University Press.
- Crystal, T.H. & House, A.S., 1988, Segmental durations in connected-speech signals : current results, *Journal of the Acoustical Society of America* 83 : 1553-1573.
- Cullicover, P. & Rochemont, M., 1983, « Stress and focus in English », *Language* 59 : 123-165.
- Dahan, D. & Bernard, J-M., 1996, Interspeaker variability in emphatic accent production in French, *Language and Speech* 39/4 : 341-374.
- De Pijper, J.R. & Sanderman, A.A., 1994, On the perceptual strength of prosodic boundaries and its relation to suprasegmental cues, *Journal of the Acoustical Society of America*, 96/4, 2037 – 2047.

- Deschamps, A., Duchet, J-L., Fournier, J-M. & O'Neil, M., 2000, *Manuel de phonologie anglaise*, Paris : Didier Erudition – CNED.
- Douglas-Cowie, E., Cowie R. & Schröder, M., 2000, A new emotion database : considerations, sources and scope, *Proceedings of the ISCA workshop on Speech and Emotion* : 39-44.
- Dubois, J., 1973, *Dictionnaire de la linguistique et des sciences du langage*, Paris : Larousse.
- Ducrot, O. & Todorov, T., 1972, *Dictionnaire Encyclopédique des Sciences du Langage*, Paris : Editions du Seuil.
- Duez, D., 1982, Silent and non-silent pauses in three speech styles, *Language and Speech*, 25 : 11-28.
- Duez, D. & Carré, R., 1979, Importance relative des paramètres de l'accent (durée et fréquence fondamentale) dans la perception de l'emphase, *International Congress of Phonetic Sciences*, Copenhagen, 374.
- Dutoit, T., 1997, *An Introduction to Text-to-Speech Synthesis*, Dordrecht : Kluwer Academic Publishers.
- Dutoit, T., Pagel, V., Pierret, N., Bataille, F. & Van Der Vrecken, O., 1996, The MBROLA project. Towards a set of high quality speech synthesizers free of use for non commercial purposes. *Proceedings ICSLP '96* (Philadelphia) 3 : 1393-1396.
- Eady, S.J., Cooper, W.E., Klouda, G.V., Mueller, P.R., Lotts, D.W., 1986, Acoustical characteristics of sentential focus : narrow vs. broad and single vs. dual focus environments, *Language and Speech* 29/3 : 233-251.
- Erickson, D. & Lehiste, I., 1995, Contrastive emphasis in elicited dialogue : durational compensation, *Proceedings of the XIIIth International Congress of Phonetic Sciences*, Stockholm 1995, 4 : 352-355.
- Fant, G., Kruckenberg, A., Liljencrants, J., 2000, The source-Filter Frame of Prominence, *Phonetica*, 57 : 113-127.
- Fant, G., Kruckenberg, A., Nord, L., 1990, Acoustic correlates of rythmical structures in text reading, in Wiik, K. & Raimo, I., (eds.), *Nordic Prosody*, n°5, University of Turku, 70-86.
- Fant, G., Kruckenberg, A. & Nord, L., 1991, Durational correlates of stress in Swedish, Fench and English, *Journal of Phonetics*, 19 : 351-365.
- Faure, G., 1962, *Recherches sur les caractères et le rôle des éléments musicaux dans la prononciation anglaise*, Paris : Didier.
- Faure, G., Hirst, D.J. & Chafcouloff, M., 1980, Rhythm in English : Isochronism, Pitch and Perceived Stress, in Waugh, L.R. & van Schooneveld, C.H., (eds.), *The Melody of Language*, Baltimore : University Park Press, 71-79.
- Firbas, J., 1972, On the interplay of prosodic and non-prosodic means of functional sentence perspective (a theoretical note on the teaching of English intonation), in

- Fried, V. (ed.), *The Prague School of Linguistics and Language Teaching*, London : Oxford University Press, 77-96.
- Fowler, C.A., 1995, Acoustic and kinematic correlates of contrastive stress accent in spoken English, in Bell-Berti, F. & Raphael, L.J., (eds.), *Producing Speech : Contemporary Issues for K. Safford Harris*, Woodbury : AIP Press, 355-374.
- Fry, D.B., 1955, Duration and intensity as physical correlates of linguistic stress, *Journal of the Acoustical Society of America* 27/4 : 765-768.
- Fry, D.B., 1958, Experiments in the perception of stress, *Language and Speech* 1 : 126-152.
- Fujisaki, H. & Nagashima, S., 1969, A model for synthesis of pitch contours of connected speech, *Annual report, Engineering Research Institute, University of Tokyo*, 28 : 53-60.
- Gimson, A.C., 1987 (3rd edition), *An Introduction to the Pronunciation of English*, London : Edward Arnold (1st edition, 1962).
- Ginésy, M., 1995, *Mémento de phonétique anglaise*, Paris : Nathan Universités.
- Glave, R.D. & Rietveld, A.C.M., 1975, Is the effort dependence of speech loudness explicable on the basis of acoustical cues ?, *Journal of the Acoustical Society of America*, 58 : 875-879.
- Goldman-Eisler, F., 1968, *Psycholinguistics : Experiments in Spontaneous Speech*, New York : Academic Press.
- Goldsmith, J.A., (ed.), 1995, *The handbook of Phonological Theory*, Cambridge, Mass. ; Oxford : Blackwell.
- Grosjean, F., 1980, Temporal variables within and between languages, in Dechert, H.W. & Raupach, M., (eds.), *Temporal Variables in Speech : Studies in Honour of Frieda Goldman-Eisler*, The Hague : Mouton, 39-53.
- Grosjean, F. & Deschamps, A., 1973, Analyse des variables temporelles du français spontané : comparaison du français oral dans la description avec l'anglais et avec le français, *Phonetica*, 28/3-4 : 191-226.
- Grosjean, F. & Deschamps, A., 1975, Analyse contrastive des variables temporelles de l'anglais et du français : vitesse de parole et variables composantes, phénomène d'hésitation, *Phonetica* 31 : 144-413.
- Groussier, M-L. & Rivière, L., 1996, *Les mots de la linguistique, lexique de linguistique énonciative*, Paris : Ophrys.
- Guaïtella I., 1995, Mélodie du geste, mimique vocale, *Semiotica*, 103-3/4 : 253-276.
- Guaïtella I., Bertrand R., Boyer J., Purson A., Espesser R., Cavé C., Santi S., Harlay F., 1998, Semiotic Value of the Link between Eyebrow Movements and Voice Variation, *The Evolution of Language, International Conference*, Londres.

- Guaïtella, I., 1991, *Rythme et Parole : Comparaison critique du rythme de la lecture oralisée et de la parole spontanée*, Thèse de Doctorat, Laboratoire Parole et Langage, Aix-en-Provence : Université de Provence.
- Gumperz, J.J., 1989, *Sociolinguistique Interactionnelle*, Saint-Denis-de-la-Réunion : l'Harmattan.
- Gumperz, J.J., 1992, Contextualization and understanding, in Duranti, A. & Goodwin, C., (eds.), *Rethinking Context. Language as an interactive phenomenon*, Cambridge : Cambridge University Press, 230-252.
- Günthner, S., 1996, The Prosodic contextualization of moral work : an analysis of reproaches in 'why'-formats, in E. Couper-Kuhlen & M. Selting, (eds.) : 271-302.
- Gussenhoven, C., 1983, « Focus, mode and the nucleus », *Journal of linguistics* 19, 377-417.
- Gussenhoven, C. & Rietveld, T., 1988, Fundamental frequency declination in Dutch : testing three hypotheses, *Journal of Phonetics*, 16 : 355-369.
- Gussenhoven, C. & Rietveld, T., 1998, On the speaker-dependence of the perceived prominence of F0 peaks, *Journal of Phonetics*, 26 : 371-380.
- Gussenhoven, C., Repp, B.H., Rietveld, A., Rump, H.H. & Terken, J., 1997, The perceptual prominence of fundamental frequency peaks, *Journal of the Acoustical Society of America*, 102/5 : 3009-3022.
- Halliday, M.A.K., 1967a, *Intonation and Grammar in British English*, Cambridge : Cambridge University Press.
- Halliday, M.A.K., 1967b, « Notes on transitivity and theme in English » part II, *Journal of Linguistics* 3 : 199-244.
- Halliday, M.A.K., 1970, *A course in spoken English : intonation*, Oxford, London : Oxford University Press.
- Handel, S., 1989, *An Introduction to the Perception of Auditory Events*, Cambridge, Mass. : MIT Press.
- Harrington, J., <<http://www.ling.mq.edu.au>>
- Hart, J. 't, 1984, A phonetic approach to intonation : from pitch contours to intonation patterns, in Gibbon, D. & Richter, H., (eds.), *Intonation, Accent and Rythm : Studies in Discourse Phonology*, Berlin, New-York : de Gruyter, 193-202.
- Hart, J. t', Collier, R. & Cohen, A., 1990, *A Perceptual Study of Intonation, an experimental-phonetic approach to speech melody*, Cambridge Studies in Speech Science and Communication, Cambridge : Cambridge University Press.
- Hartmann, R. & Stork, F., 1972, *Dictionary of Language and Linguistics*, London : Applied Science Publications.
- Haslerud, V. & Stenstrom, A-B., 1995, The Bergen Corpus of London Teenage (COLT), in Leech, G., Meyres, G & Thomas, J., (eds), *Spoken English on Computer*, London : Longman, 235-242.

- Hayes, B., 1994, « Gesture » in prosody : comments on the paper by Ladd, in Keating, P.A., (ed.), *Phonological Structure and Phonetic Form. Papers in Laboratory Phonology III*, Cambridge : Cambridge University Press, 64-75.
- Herment-Dujardin, S., 1997, *Experimental validation of the intonation transcription system INTSINT in English*, Mémoire de DEA, Aix-en-Provence : Université de Provence.
- Hirschberg, J. & Ward, G., 1992, The influence of pitch range, duration, amplitude and spectral features in the interpretation of rise-fall-rise intonation contour in English, *Journal of Phonetics* 20 : 241-251.
- Hirst, D.J., 1977a, Emphatic intonation in generative grammar, in Carton, F., Hirst, D.J., Marshall, A. & Séguinot, A., (eds.), *L'accent d'insistance : emphatic stress*, Paris : Studia Phonetica, Didier, 12 : 123-143.
- Hirst, D.J., 1977b, *Intonative Features : a Syntactic Approach to Prosody*, The Hague : Mouton.
- Hirst, D.J., 1987, *La représentation linguistique des systèmes prosodiques : une approche cognitive*, Thèse de Doctorat d'Etat, Institut de Phonétique d'Aix-en-Provence : Université de Provence.
- Hirst, D.J., 1998, Intonation in British English, in D.J. Hirst & A. Di Cristo, (eds.) : 56-77.
- Hirst, D.J., 1999, The symbolic coding of duration and timing : an extension to the INTSINT system, *Proceedings Eurospeech '99* (septembre 1999, Budapest), 1639-1642.
- Hirst, D.J., 2000, ProZed: a multilingual prosody editor for speech synthesis, *Proceedings, IEE Workshop State of the Art in Speech Synthesis*, London, March 2000 : 4/1-4/7.
- Hirst, D.J. & Di Cristo, A., 1998, *Intonation Systems : a Survey of Twenty Languages*, Cambridge : Cambridge University Press.
- Hirst, D.J., Di Cristo, A. & Espesser, R. 2000, Levels of representation and levels of analysis for intonation, in M. Horne, (ed.) : 51-88.
- Hirst, D.J., Di Cristo, A., Le Besnerais, M., Najim, P., Roméas, N.P., 1993, Multilingual modelling of intonation patterns, *Proceedings ESCA Workshop on Prosody*, Lund, Septembre 1993, 204-207.
- Hirst, D.J., Espesser, R., 1993, Automatic modelling of fundamental frequency using a quadratic spline function, *Travaux de l'Institut de Phonétique d'Aix-en-Provence*, 15 : 75-85.
- Hollien, H., Hollien, P.A., de Jong, G., 1997, Effects of three parameters on speaking fundamental frequency, *Journal of the Acoustical Society of America*, 102/5 : 2984-2992.
- Horne, M., (ed.), 2000, *Prosody : Theory and Experiment*, Dordrecht : Kluwer Academic Publishers.

- House, A., 1961, On vowel duration in English, *Journal of the Acoustical Society of America* 33, 174-178.
- Householder, F., 1957, Accent, Juncture, Intonation, and My Grandfather's Reader, *Word* 13 : 234-245.
- IViE, voir <<http://www.phon.ox.ac.uk/~esther/ivyweb>>
- Jackendoff, R.S., 1972, *Semantic Interpretation in Generative Grammar*, Cambridge, Mass. : MIT Press.
- Jassem, W., 1952, *Intonation of Conversational English (Educated Southern British)*, Wrocław : Wrocławskie Towarzystwo Naukowe.
- Jones, D., 1918, *Outline of English Phonetics*, 8th edition : 1956, Cambridge : Heffer.
- Kim, H. & Loh, W-Y., sous presse, Classification trees with unbiased multiway splits, *Journal of the American Statistical Association*, 96.
- Kim, H. & Loh, W-Y., 2001, CRUISE, User Manual, *Technical Report 989*, March 3, 1998, revised November 10, 2001, Department of Statistics, University of Wisconsin, Madison.
- Kingdon, R., 1958a, *The Groundwork of English Stress*, London : Longman.
- Kingdon, R., 1958b, *The Groundwork of English Intonation*, London : Longman.
- Klatt, D.H., 1975, Vowel lengthening is syntactically determined in a connected discourse, *Journal of Phonetics* 3, 129-140.
- Klatt, D.H., 1976, Linguistic uses of segmental duration in English : Acoustic and perceptual evidence, *Journal of the Acoustical Society of America* 59 : 1208-1221.
- Klatt, D.H., 1979, Synthesis by rule of segmental durations in English sentences, in Lindblom, B. & Öhman, S. (eds.), *Frontiers of Speech Communications Research*, New York : Academic Press, 287-299.
- Klatt, D.H., 1987, Review of text-to-speech conversion for English, *Journal of the Acoustical Society of America* 82 : 737-793.
- Kohler, K.J., 1997, Modelling Prosody in Spontaneous Speech, in Y. Sagisaka *et al.*, (eds.) : 187-210.
- Ladd, R., 1980, *The structure of intonational meaning : evidence from English*, Bloomington : Indiana University Press.
- Ladd, R., 1984, Declination : a review and some hypotheses, in Ewen, C.J. & Anderson, J.M., (eds), *Phonology Yearbook 1*, Cambridge : Cambridge University Press, 53-74.
- Ladd, R., 1990, « Intonation : emotion versus grammar », *Language* 66/4 : 804-816.
- Ladd, R., 1993, On the theoretical status of the 'baseline' in modelling intonation, *Language and Speech*, 36/4 : 435-451.

- Ladd, R., 1994, Constraints on the gradient variability of pitch range, or, pitch level 4 lives !, in Keating, P.A., (eds.), *Phonological Structure and Phonetic Form. Papers in Laboratory Phonology III*, Cambridge : Cambridge University Press, 43-63.
- Ladd, R. & Morton, R., 1997, The perception of intonational emphasis : continuous or categorical ?, *Journal of Phonetics*, 25 : 313-342.
- Ladd, R., Scherer, K.R., Silverman, K.E.A., 1986, An integrated approach to studying intonation and attitude, in Johns-Lewis, C., (ed.), *Intonation in Discourse*, London & Sydney : Crook Helm, 125-138.
- Ladd, R., Silverman, K.E.A., Tolkmitt, F., Bergmann, G. & Scherer, K.R., 1985, Evidence for the independent function of intonation contour type, voice quality, and F0 range in signalling speaker affect, *Journal of the Acoustical Society of America*, 78/2 : 435-444.
- Ladd, R., Verhoeven, J. & Jacobs, K., 1994, Influence of adjacent pitch accents on each other's perceived prominence : two contradictory effects, *Journal of Phonetics*, 22 : 87-99.
- Ladefoged, P., 2001, *Vowels and Consonants : an Introduction to the Sounds of Languages*, Oxford : Blackwell.
- Lambrecht, K., 1994, *Information Structure and Sentence Form*, Cambridge : Cambridge University Press.
- Lehiste, I., 1963, Review of K. Hadding-Koch, Acoustico-phonetic studies in the intonation of Southern Swedish, *Glerup Language*, 39 : 352-360.
- Lehiste, I., 1977, Isochrony reconsidered, *Journal of Phonetics* 5, 253-263.
- Liberman, M. & Pierrehumbert, J., 1984, Intonational invariance under changes in pitch range and length, in Aranof, M. & Oerhle, R. (eds.), *Language Sound Structure : Studies in Phonology Presented to Morris Halle*, Cambridge, Mass. : MIT Press, 157-233.
- Liberman, M., 1975, *The Intonational System of English*, PhD Dissertation, Cambridge, Mass. : MIT.
- Lieberman, P., 1967, Intonation, perception and Language, *Research Monograph*, 38, Cambridge, Mass. : MIT Press.
- Lieberman, P., 1974, A study of prosodic features, in Sebeok, Th.A. (ed.), *Current Trends in Linguistics*, Vol. 12, The Hague : Mouton, 2419-2449.
- Lieberman, P. & Michaels, S.D., 1962, Some aspects of fundamental frequency, envelope amplitude and the emotional content of speech, *Journal of the Acoustical Society of America* 34 : 922-927.
- Lilly, R. & Viel, M., 1999 (édition révisée), *Initiation raisonnée à la phonétique de l'anglais*, Paris : Hachette Supérieur.

- Louis, M., en cours, *Contribution à l'étude expérimentale des pathologies dégénératives du langage et de la parole*, Thèse de Doctorat, Laboratoire Parole et Langage, Aix-en-Provence : Université de Provence.
- Maeda, S., 1976, *A characterization of American English Intonation*, Doctoral Dissertation, MIT, Cambridge.
- Malfrère, F. & Dutoit, T., 1997, High Quality Speech Synthesis for Phonetic Speech Segmentation, *Proceedings Eurospeech '97* : 2631-2634.
- MARSEC, <<http://www.midwich.reading.ac.uk/research/speechlab/marsec.html>> ou <<http://www.khnt.hit.uib.no/icame/sec/INDEX.HTM>>
- Massaro, D.W., 1987, *Speech perception by ear and eye : a paradigm for psychological inquiry*, Hillsdale, NJ : Lawrence Erlbaum Associates.
- MBROLA, disponible sur <<http://tcts.fpms.ac.be/synthesis/mbrola.html>>
- Morton, R., 1993, *On the structure of pitch range in intonation : gradient or categorical ?*, Honours Dissertation, University of Edinburgh.
- Mounin, G., 1974, *Dictionnaire de la Linguistique*, Paris : PUF.
- Mozziconacci, S., 1998, *Speech Variability and Emotion : Production and Perception*, Eindhoven : Technische Universiteit Eindhoven.
- Nespor, M., and Vogel, I., 1986, *Prosodic Phonology*, Dordrecht : Foris.
- Ní Chasaide, A. & Gobl, C., 2000, Testing affective correlates of voice quality through analysis and resynthesis, *Proceedings of the ISCA workshop on Speech and Emotion* : 178-183.
- Nicaise, A. & Gray, M., *L'intonation de l'anglais*, Paris : Nathan, Collection « 128 ».
- O'Connor, J.D., 1965, The perception of time intervals, *Progress Report 2*, Phonetics Laboratory, University College, London, 11-15.
- O'Connor, J.D., Arnold, G.F., 1973, *Intonation of Colloquial English*, London : Longman.
- O'Shaughnessy, D., 1979, Linguistic features in fundamental frequency patterns, *Journal of Phonetics*, 7 : 119-145.
- Ohala, J. & Hirano, M., 1967, An experimental investigation of pitch change in speech (abstract), *Journal of the Acoustical Society of America*, 42 : 1208-1209.
- Osgood, C.E., Suci, G.J. & Tannenbaum, P.H., 1957, *The Measurement of Meaning*, Urbana, Illinois.
- Osser, H., 1964, *A 'Distinctive Features' Analysis of the Vocal Communication of Emotion*, Dissertation, Cornell University.
- Pagel, V., Carbonell, N., Laprie, Y., 1996, A new method for speech delexicalisation and its application to the perception of French prosody, *Proceedings ICSLP'96* (Philadelphia) : 821-824.

- Passy, P.E., 1906, *Les sons du français, leur formation, leur combinaison, leur représentation*, Paris : Firmin-Didot.
- Peeters, J., 1999, Thématization et focalisation : deux principes distincts de construction du sens, in Guimier, C. (ed.), *La thématization dans les langues, Actes du colloque de Caen*, 9-10 octobre 1997, Bern : Peter Lang, 45-61.
- Peterson, G.E. & Lehiste, I., 1960, Duration of syllable nuclei in English, *Journal of the Acoustical Society of America* 32, 693-703.
- Pierrehumbert, J. & Hirschberg, J., 1990, The Meaning of Intonation Contours in the Interpretation of Discourse, in Cohen, P.R., Morgan, J. & Pollack, M.E., (eds.), *Intentions in Communication*, Cambridge, Mass. : MIT Press, 271-311.
- Pierrehumbert, J. & Talkin, D., 1992, Lenition of /h/ and glottal stop, in Docherty, G. & Ladd, D.R., (eds.), *Papers in Laboratory Phonology II : Gesture, Segment, Prosody*, Cambridge : Cambridge University Press, 90-116.
- Pierrehumbert, J., 1979, The perception of fundamental frequency declination, *Journal of the Acoustical Society of America*, 66 : 363-369.
- Pierrehumbert, J., 1980, *The phonology and phonetics of English intonation*, PhD Dissertation, MIT.
- Pitrelli, J., 1990, *Hierarchical Modeling of Phoneme Duration Applied to Speech Recognition*, PhD Thesis, MIT.
- PRAAT, <<http://www.fon.hum.uva.nl/praat>> ou <<http://www.praat.org>>
- Purson A., Santi S., Bertrand R., Guaitella I., Boyer J., Cavé C. (1999) The Relationships between Voice and Gesture : Eyebrows Movements and Questioning, *VIth European Conference on Speech Communication and Technology*, Budapest (soumis).
- Quirk R., Greenbaum, S., Leech, G. & Svartvik, J., 1985, *A Comprehensive Grammar of the English Language*, London and New York : Longman.
- Quirk, R., Greenbaum, S., Leech, G. & Svartvik, J., 1972, *A Grammar of Contemporary English*, London : Longman.
- Quirk, R., Duckworth, A. P., Svartvik, J., Rusiecki, J.P.L. & Colin, A.J.T., 1964, Studies in the correspondance of prosodic to grammatical features in English, *Proceedings of the IX international Congress of Linguistics*, The Hague : Mouton, 679-691.
- Ramus, F. & Mehler, J., 1999, Language identification with suprasegmental cues : a study based on speech resynthesis, *Journal of the Acoustical Society of America*, 105 : 512-521.
- Repp, B., 1992, Probing the cognitive representation of musical time : structural constraints on the perception of timing perturbations, *Haskins Laboratories Status Report on Speech research*, 111/112, New Haven, Connecticut, 293-320.

- Riley, M., 1992, Tree-based modelling of segmental durations, in Bailly, G. & Benoît, C. (eds.), *Talking Machines*, Amsterdam : North-Holland, 265-273.
- Roach, P., 1983 (third edition, 2000), *English Phonetics and Phonology, a practical course*, Cambridge : Cambridge University Press.
- Roach, P., 1994, Conversion between prosodic transcription systems : « Standard British » and ToBI, *Speech Communication* 15 : 91-99.
- Roach, P., 2000, Techniques for the description of emotional speech, *Proceedings of the ISCA workshop on Speech and Emotion* : 53-59.
- Roach, P., Knowles, G., Varadi, T., Arnfield, S., 1993, MARSEC : A Machine-Readable Spoken English Corpus, *Journal of the International Phonetic Association*, 23/2 : 47-53.
- Rouveyrol, L., en cours, *Etude pragmatique de la variation linguistique dans le débat politique médiatisé en anglais. Modalités et marques d'implication du locuteur dans le discours. Vers une stylistique de l'interaction télévisée*, Thèse de Doctorat, Laboratoire Parole et Langage, Aix-en-Provence : Université de Provence.
- Rump, H.H. & Collier, R., 1996, Focus Conditions and the Prominence of Pitch-Accented Syllables, *Language and Speech*, 39/1 : 1-17.
- Sagisaka, Y., Campbell, N., Higuchi, N., (eds.), 1997, *Computing Prosody*, New York : Springer-Verlag.
- SAMPA, <<http://www.phon.ucl.ac.uk/home/sampa/x-sampa.htm>>
- Schmerling, S., 1976, *Aspect of English Sentence Stress*, Austin : Texas University Press.
- Schmerling, S., 1980, The proper treatment of the relationship between syntax and phonology. *Paper presented at the fifty-fifth annual meeting of the Linguistic Society of America*, December 1980, San Antonio, Texas.
- Selkirk, E., 1984, *Phonology and Syntax*. Cambridge, Mass. : MIT Press.
- Selkirk, E., 1995, Sentence Prosody : Intonation, Stress and Phrasing, in J.A. Goldsmith, (ed.) : 550-569.
- Selting, M., 1989, Speech Styles in conversation as an interactive achievement, in Hickey, L., (ed.), 1989, *The Pragmatics of Style*, London and New York : Routledge, 106-132.
- Selting, M., 1994, Emphatic speech style - with special focus on the prosodic signalling of heightened emotive involvement in conversation, *Journal of Pragmatics*, 22 : 375-408.
- Selting, M., 1996, Prosody as an activity-type distinctive cue in conversation : the case of so-called 'astonished' questions in repair initiation, in E. Couper-Kuhlen & M. Selting, (eds.) : 231-270.
- Shen, Y. & Peterson, G., 1964, Isochronism in English, *Occasional Papers* 9, Studies in Linguistics, Buffalo : University Press.

- Silverman, K., Beckman, M., Pitrelli, J., Ostendorf, M., Wightman, C., Price, P., Pierrehumbert, J. & Hirschberg, J., 1992, ToBI : a standard for labelling English Prosody, *Proceedings ICSLP '92*, 2 (Banf, Canada) : 867-870.
- Sluitjer, A., 1995, Phonetic Correlates of Stress and Accent, *Holland Institute of General Linguistics*.
- Sluitjer, A. & van Heuven, V., 1993, Perceptual clues of linguistic stress : intensity revisited, *Working Papers 41, Proceedings of the ESCA Workshop on Prosody*, Lund University, Sweden, 246-249.
- Sluitjer, A. & van Heuven, V., 1996, Spectral balance as an acoustic correlate of linguistic stress, *Journal of the Acoustical Society of America* 100/4 : 2471-2485.
- SproSIG : home page : <<http://www.lpl.univ-aix.fr/projects/sprosig>> ; archives : <<http://www.egroups.com/group/sprosig>>
- Steedman, M. J., 1990, Syntax and Intonational Structure in a Combinatory Grammar, in Altmann, G., (ed.), *Cognitive Models of Speech Processing, Psycholinguistic and Computational Perspectives*, Cambridge, Mass. : MIT Press, 457-482.
- Steedman, Mark, 1991. Structure and Intonation, *Language*, 67 : 260-296.
- Streefkerk, B., 1997, Acoustical correlates of prominence : a design for research, *Proceedings of the Institute of Phonetic Sciences of the University of Amsterdam*, 21 : 131-142.
- Streefkerk, B.M., Pols, L.C.W. & ten Bosh, L.F.M., 1997, Prominence in read aloud sentences, as marked by listeners and classified automatically, *Proceedings of the Institute of Phonetic Sciences of the University of Amsterdam*, 21 : 101-116.
- Svartvik, J. & Quirk, R., (eds.), 1980, A corpus of English Conversation, *Lund Studies in English* 56, Lund : Liber/Gleerups.
- Svartvik, J., (ed.), 1990, The London Corpus of Spoken English, Description and Research, *Lund Studies in English* 82, Lund : Lund University Press.
- Sweet, H., 1906, *A primer in Phonetics*, Oxford : Clarendon.
- Tannen D., 1984, *Conversational Style*, Norwood, NJ : Ablex.
- Tartter, V.C. & Braun D., 1994, Hearing smiles and frowns in normal and whisper registers, *Journal of the Acoustical Society of America*, 96/4 : 2101-2107.
- Taylor, L.J. & Knowles, G., 1988, Manual of information to accompany the Spoken English Corpus, *Technical Report, UCREL*, University of Lancaster.
- Terken, J., 1991, Fundamental frequency and perceived prominence of accented syllables, *Journal of the Acoustical Society of America* 89/1 : 1768-1776.
- Terken, J., 1993, Baselines revisited : reply to Ladd, *Language and Speech*, 36/4 : 453-459.
- Terken, J., 1997, Variation of Accent Prominence within the Phrase : Models and Spontaneous Speech Data, in Y. Sagisaka *et al.*, (eds.) : 95-116.

- Thorsen, N., 1980, A study of the perception of sentence intonation – evidence from Danish, *Journal of the Acoustical Society of America*, 67 : 1014-1030.
- Trager, G.L. & Smith, H.L., 1951, *An Outline of English Structure*, Norman, Oklahoma : Battenburg Press.
- Trask, R.L., 1993, *A Dictionary of Technical Terms in Linguistics*, London and New York : Routledge.
- Trouvain, J., et Barry, W.J., 2000, The Prosody of excitement in horse race commentaries, *Proceedings of the ISCA workshop on Speech and Emotion* : 86-91.
- Uhmman, S., 1992, Contextualizing relevance : on some forms and functions of speech rate changes in everyday conversation, in P. Auer & A. di Luzio (eds.) : 297-336.
- Uhmman, S., 1996, On rythm in everyday German conversation : beat clashes in assessment utterances, in E. Couper-Kuhlen & M. Selting, (eds.) : 303-365.
- Uldall, E., 1964, Dimensions of meaning in intonation, in D. Abercrombie *et al.*, (eds.) : 271-279.
- Vaissière, J., 1983, Language-independent prosodic feature, in Cutler A. & Ladd, R., (eds.), *Prosody : Models and measurements*, Heidelberg : Springer : 53–66.
- Ward, G. & Hirschberg, J., 1985, Implicating uncertainty : the pragmatics of fall-rise intonation, *Language* 61 : 747-776.
- Ward, G. & Hirschberg, J., 1988, Intonation and propositional attitude : the pragmatics of L*+HLH%, *Proceedings of the fifth meeting of the Eastern states conference on linguistics*, Philadelphia.
- Weismer, G. & Ingrisano D., 1979, Phrase-level timing patterns in English : the effects of emphatic stress location and speaking rate, *Journal of Speech and Hearing*, 22 : 516-533.
- Wells, J.C., 1995, Computer-coding of the IPA : a proposed extension of SAMPA. ms.
- Wells, J.C., Barry, W., Grice, M., Fourcin, A., Gibbon, D., 1992, Standard computer-compatible transcription, *Esprit project 2589 (SAM)*, Doc. no. SAM-UCL-037, London, Phonetics and Linguistics Department, UCL.
- Wells, W.H.G., 1986, An experimental approach to the interpretation of focus in spoken English, in Johns-Lewis, C., (ed.), *Intonation in Discourse*, London & Sydney : Crook Helm : 53-75.
- Winkler, P., 1984, Interrelations between Fundamental Frequency and Other Acoustic Parameters of Emphatic Segments, in Gibbon, D. & Richter, H., (eds), *Intonation, Accent and Rythm*, Studies in Discourse Phonology : Research in Text Theory, Berlin ; New York : de Gruyter, 327-338.
- Witten, I.H., 1977, A flexible scheme for assigning timing and pitch to synthetic speech, *Language and Speech* 20 : 240-260.

- Young, S.D. & Fallside, F., 1980, Synthesis-by-rule of prosodic features in word concatenation systems, *International Journal of Man-Machine Studies* 12 : 241-258.

Index

Index des notions et mots-clés

A

accent (voir aussi désaccentuation, default
accent)
contrastif 32,34,57,58,178
d'intensité.....55
de phrase36
lexical
36,37,38,48,53,55,112,137,148,153,252,
260
mélodique..... 37,48,49,5182,198
musical36
nucléaire..... 26,27,28,29,30,40,49
prénucléaire..... 30,33,40,50
secondaire38
silencieux38
amplitude totale 197,223,224,225
attaque179,181,253,257,260,266,281

B

BNC72

C

COLT72,290
contraste 19,22,24,31,32,33,34,39,41,54,55,65
CRUISE 14,185ff,277,278,292
CSAE72

D

déclinaison 50,59,82,247
default accent30

délexicalisation

..... 237,238,248,265,269,272,273,280,282
aaaa 248
jastadanz.....249,262,267,271,315
saltanaj..... 248,249
sasasa 248
désaccentuation 23,30

E

émotion

..17,20,21,22,41,42,43,44,45,46,47,48,70,182

emphase

contrastive32,33,54,55,61,65
d'intensité..... 32,33,55,65
sens rhétorique 18,40

F

focalisation 19,20,22,23,30,64,71,295
focus 23,24,25
contrastive 56
étroit25,27,28,29,30,31,40,59
large.....25,28,29,31,40,59
neutre 25,59
non marqué..... 28
portée du..... 29,31,53,59
principal..... 56
subsidaire 56
zéro 56

G

gestuelle..... 21,22,48

glissement accentuel..... 34,38,112
glotte.....120,178,179,181,182,253,281

I

ICE-GB72

implication

..... 18,21,22,40,45,47,60,67,84,276,281,284

information

ancienne23,24,25,26,27,30,40,88

contrastive.....31

nouvelle..... 19,22,23,24,25,26,30,31,32,40,88

IPA 106,110,242,298,315

isochronie.....38

IViE.....73,292

L

LLC72

M

MARSEC 73,294,296

MBROLA 14,242ff,278,288,294

MOMEL..... 123,124,140,141

mouvement mélodique

dive..... 144,145,239

glide.....140

high jump241

ondulatoire 140,141,146,239,262

P

PERL.....106,107,108,151,249,315

piéd..... 132,133,134,136,137,149

PRAAT 14,95ff,242,244,277,285,295,315

pragmatique13,22,24,27,34,65,161,236,284

présupposition..... 23,24,25,32

proéminence

22,23,35,36,37,38,39,41,48,49,50,51,52,54,

56,66,80,83,84,118,131,148,152,267

prosodie 12,20,22,23,46,60,61,237,280,282

Q

qualité de la voix

..... 43,46,47,67,83,122,126,162,182

R

resynthèse vocale 237

rhème..... 22,23

rythme.....38,61,67,89,248,290

S

SAMPA .. 106,108,110,242,243,244,296,298,315

script107,108,146,151,157,278

SEC..... 72

sémantique

18,22,23,24,26,28,29,32,34,37,40,43,55,111,

143,237,255,266,268,272,273,280,281,282

SEU 72

SproSIG..... 61,297

SSE 72

syntaxe..... 19,22,29,240,267,269,272,273,282

T

thématisation 23,24,295

thème 23,24

tilt spectral.....52,53,67,278,286,297

ton

circonflexe..... 136,193

creusé.....81,135,139,193,238,244

topicalisation 22,24

topique..... 19,22,23,24,26,39

U

unité

intonative

33,40,60,86,133,137,138,140,141,142,14

3,146,148,157,162,164,177,183,224,225,
239,240,252,254,263,279

tonale

132,133,134,138,139,141,142,146,156,1

57

*Index des auteurs***A**

Abercrombie 38,133,134,284,298
 Armstrong 143,284
 Arnold 45,140,285,287,289,294
 Arrivé 19,20,284
 Asher 18,284
 Astésano 123,284
 Auran 106,151,284

B

Bagshaw 49,284
 Barnwell 149,284
 Barry 46,285,298
 Beckman
 14,39,49,50,52,54,67,71,153,197,198,279,
 280,285,286,297
 Bernard 64,65,287
 Blanchard 185,285
 Bloomfield 22,36,285
 Boersma 14,95,104,118,124,125,153,277,285
 Bolinger 21,26,28,33,34,35,37,45,47,285
 Bouscaren 23,24,285
 Bouzon 106,185,285
 Braun 48,297
 Bright 18,285
 Brown 31,32,49,143,285
 Bussman 18,21,40,45,285

C

Campbell
 35,52,54,67,106,148,149,150,151,153,180,
 246,279,286,296
 Cavé 48,286,289,295
 Chafe 22,23,72,286
 Chomsky 22,23,25,28,286
 Chuquet 285
 Classe 21,38,214,286
 Clear 72,287
 Cohen 13,290,295
 Coker 149,287
 Coleman 32,54,55,63,67,287
 Collier 13,290,296
 Collinge 18,287
 Cooper 50,57,59,82,287,288
 Couper-Kuhlen
 26,31,32,38,45,144,287,290,296,298
 Cowie 73,288
 Croft 287
 Cruttenden 27,28,29,30,31,143,144,145,287
 Crystal 20,21,37,38,45,143,149,287

D

Dahan 64,65,287
 Danon-Boileau 285
 Deschamps 144,160,288,289
 Di Cristo 24,31,32,64,67,137,282,291

Douglas-Cowie73,288
 Dubois 19,20,286,288
 Duchet144,288
 Ducrot19,288
 Dutoit 14,242,278,288,294

E

Eady 57,59,287,288
 Edwards..... 49,50,285
 Erickson 61,67,288
 Espesser.....123,137,282,284,286,289,291

F

Fallside149,299
 Faure.....81,288
 Firbas.....23,288
 Fournier.....144,288
 Fry 48,49,67,284,289
 Fujisaki.....51,289

G

Gadet284
 Galmiche284
 Gimson.....118,289
 Ginésy 142,144,145,289
 Glave289
 Gobl.....182,294
 Goldman-Eisler..... 159,161,171,289
 Grabe73
 Gray 144,145,294
 Grosjean..... 159,160,289
 Groussier 19,20,289
 Guaïtella.....48,160,286,289,290,295

Gumperz 290
 Günthner..... 41,290
 Gussenhoven 22,24,51,66,80,82,143,248,290

H

Halle 28,286,293
 Halliday22,23,25,45,134,143,290
 Harrington 50,290
 Hartmann..... 20,290
 Haslerud 72,290
 Hayes 35,291
 Herment-Dujardin..... 123,291
 Higuchi 296
 Hirano..... 143,294
 Hirschberg47,80,291,295,297,298
 Hirst
 24,30,31,32,33,64,67,107,108,123,133,137,
 138,140,143,146,282,284,288,291
 House.....137,138,149,238,287,292
 Householder 38,292

I

Ingrisano..... 49,298
 Isard..... 150,286

J

Jackendoff 22,23,25,292
 Jassem..... 133,134,143,292
 Jones 55,284,292

K

Kim..... 14,185,187,198,202,277,292
 Kingdon..... 36,143,292
 Klatt 149,292

Klouda.....59,288
Knowles296,297

L

Ladd
22,23,25,27,28,30,38,44,51,79,80,81,82,83,
209,214,278,291,292,293,295,297,298
Lambrecht24,293
Lehiste.....38,61,67,71,148,288,293,295
Lieberman 13,38,50,293
Lieberman 42,44,63,143,293
Lilly 179,181,293
Llisterri.....284
Loh14,185,187,198,202,277,292

M

Maeda.....50,294
Massaro48,294
McGlone49,285
Mehler 236,248,249,295
Morton.....79,80,81,82,83,293,294
Mounin 19,20,294
Mozziconacci46,294
Mueller 57,59,287,288

N

Nagashima51,289
Nespor137,294
Ní Chasaide.....182,294
Nicaise..... 144,145,294
Nolan.....73

O

Ohala143,294

Osgood 43,294
Osser 43,44,294

P

Pagel236,237,288,294
Passy 20,36,295
Peeters 23,24,295
Peterson 38,148,295,296
Pierrehumbert..... 13,47,50,51,82,293,295,297
Pitrelli 149,295,297
Pols 83,297
Purson.....48,289,295

Q

Quirk..... 21,22,23,24,72,143,295,297

R

Ramus236,248,249,295
Repp 160,161,290,295
Rietveld51,66,80,82,248,289,290
Riley 149,296
Rivière 19,20,289
Roach..... 73,296
Rouveyrol..... 74,296
Rump 290,296

S

Sagisaka.....285,286,292,296,297
Scherer..... 293
Schmerling22,23,26,28,296
Schröder 73,288
Selkirk 137,296

Selting		Trask.....	19,20,298
13,21,40,44,45,52,61,62,67,140,287,290,296		Trouvain	46,298
,298		U	
Shen.....	38,296	Uhmann	13,52,60,61,67,140,298
Silverman	293,297	Uldall.....	43,44,46,298
Sluitjer.....	52,53,54,67,152,153,248,279,297	V	
Smith	43,63,298	Vaissière.....	51,298
Sorensen.....	50,287	van Heuven	52,53,54,152,153,248,279,297
Steedman.....	297	Viel	179,181,293
Stenstrom	72,290	Vogel	137,294
Stork	20,290	W	
Streefkerk.....	83,84,131,297	Ward	47,80,143,284,291,298
Svartvik	72,295,297	Weenik	14,95,118,124,153,277,285
Sweet.....	36,143,297	Weismer	49,298
T		Wells, J.C.	106
Talkin	50,295	Wells, W.H.G.....	56
Tannen.....	46,60,67,287,297	Winkler.....	62,63,298
Tartter.....	48,297	Witten	149,298
ten Bosch.....	83	Y	
Terken	35,49,82,290,297	Young	149,299
Thorsen	51,298		
Todorov.....	19,288		
Trager	43,63,298		

Table des annexes

Table des annexes.....	308
Annexe 1 - IPA / SAMPA	309
Annexe 2a - Sondages : auditeurs	310
Annexe 2b - Sondages : instructions	311
Annexe 2c - Script 'TextGrid2Matrix'	335
Annexe 2d - Script 'AddPron2Matrix'	337
Annexe 3a - Script 'measure F0'	340
Annexe 3b - 'measure intensity'	342
Annexe 3c - Script 'mtostat'	344
Annexe 3d - Intensité : pré analyses.....	348
Annexe 3e - Nombre de pauses.....	353
Annexe 3f - Durée des pauses.....	354
Annexe 3g - Pauses emphatiques, avant ou après.....	355
Annexe 4a - Liste des abréviations pour les variables utilisées avec CRUISE	356
Annexe 4b - Arbres de classification : corpus primaire	359
Annexe 4c - Arbres de classification : types de discours	370
Annexe 4d - Arbres de classification : locuteurs.....	372
Annexe 5a - Script 'mapPhones'	378
Annexe 5b - Test perceptif sur signaux manipulés : auditeurs	380
Annexe 5c - Test perceptif sur signaux manipulés : instructions.....	381
Annexe 5d - Test perceptif sur signaux manipulés : résultats.....	386
Annexe 5e - Test perceptif sur signaux manipulés : phrases-sources	390
Contenu du CD ROM	391

Annexe 1 - IPA / SAMPA

Correspondance entre l'alphabet phonétique international (IPA) et l'alphabet SAMPA

IPA	SAMPA
I	I
i :	i:
e	E
{	{
A :	A:
	Q
O :	O:
U	U
u :	u:
	V
:	3
@	@
aI	aI
eI	eI
OI	OI
aU	aU
@U	@U
I@	I@
U@	U@
e@	e@
aI@	aI@
eI@	eI@
OI@	OI@
AU@	aU@
@U@	@U@

IPA	SAMPA
p	p
b	b
t	t
d	d
k	k
g	g
f	f
v	v
s	s
z	z
S	S
Z	Z
tS	tS
dZ	dZ
T	T
D	D
m	m
n	n
	N
r	r
w	w
j	j
l	l
	5
h	h

Annexe 2a - Sondages : auditeurs

Sondage 1 :

A : femme 25 ans anglaise lectrice à Aix vit en France depuis 3 ans
 AD : femme 60 ans Anglaise vit en France depuis très longtemps ; MCF anglais des affaires fac lettres Aix
 AH : femme, 30/35 ans Américaine ATER Civi US fac lettres Aix
 CB : femme 23 ans anglaise étudiante
 CC homme 22 ans irlandais lecteur
 CP : femme 35 ans anglaise MCF litté trad fac lettres Aix
 CR homme 38 ans anglais ancien lecteur pho
 DV : homme 26 ans anglais lecteur
 DT : homme Américain 30 ans ATER Lit. US fac lettres Aix
 GP : femme 33 ans Américaine Etudiante en 2ème année anglais à Aix
 JH homme 30 ans anglais lecteur fac lettres Aix
 MG. : homme 40 ans anglais MCF civ brit Fac Lettres Aix
 MH homme 50 ans anglais MCF litté trad fac lettres Aix
 PP : homme 45 ans anglais MCF ling fac lettres Aix
 RT : homme 50 ans Anglais MCF linguistique fac de lettres Aix
 SN femme 24 ans anglaise ancienne lectrice
 SO : femme 27 ans Ecossaise vit en France depuis 6 ans ATER civ brit fac lettres Aix
 TB : femme 20 ans anglaise lectrice
 MC: femme 44 ans anglaise mère au foyer

Sondages 2 & 3 :

AD : femme 60 ans Angl vit en France depuis très longtemps ; MCF anglais des affaires fac lettres Aix
 AH : femme, 34 ans Américaine ATER Civi US fac lettres Aix
 CB : femme 23 ans anglaise étudiante
 CC : homme 23 ans irlandais lecteur
 CP : femme 35 ans ? anglaise MCF litté trad fac lettres Aix
 DV : homme 26 ans anglais lecteur
 GP : femme 33 ans Américaine Etudiante en 2ème année anglais à Aix
 JH homme 30 ans anglais lecteur fac lettres Aix
 MG : homme 40 ans anglais MCF civ brit Fac Lettres Aix
 MW : femme 75 ans canadienne retraitée
 PP : homme 45 ans anglais MCF ling fac lettres Aix
 R T : homme 50 ans Anglais MCF linguistique fac de lettres Aix
 RE : homme 28 ans américain lecteur
 RF : femme 30 ans anglaise prof d'anglais en France
 RS : homme 38 ans prof d'anglais en Italie
 TB : femme 20 ans anglaise lectrice
 MC : femme 44 ans anglaise mère au foyer

Annexe 2b - Sondages : instructions

Sondage 1 : Question Time

Dear colleagues,

For my research work on **emphasis** in English, I need to do a little poll among native speakers. What you have to do is listen to a TV debate and underline all the passages or words which you feel are **EMPHATIC**. What I mean when I say emphatic is NOT NEUTRAL, with a special INVOLVEMENT on the part of the speaker. For example, ask yourselves if you would utter the same sentence in the same way in a neutral context. What I want you to tell me is what you think, what you feel is being made prominent in some way.

I am not giving you a precise definition of emphasis because this is what I intend to find once I have the results of this test.

So if you could spend some time (about 30 mns I think) listening to the tape (side A only) and filling in the paper that goes with it, I would be very grateful.

You are first going to listen to models : a sentence uttered in a neutral way and the same sentence uttered in an emphatic way. This is to let you hear what I mean by emphasis. Then you'll hear a debate (it is the TV programme *Question Time* in Belfast here). You first hear a short passage twice and then I have selected sentences which you will hear twice and which are written in italics on the script. I would like you to underline on the text in italics the word(s) which you perceive as being emphatic. I have left some space for you to make comments which might be useful to me. For instance you can tell me if you hesitated or had doubts or felt the emphasis was less strong or if you thought it was ambiguous or whatever might help me. Of course, not all the passages you are going to hear are emphatic.

You can reach me on 04 42 23 16 42 if you feel the need. I would also be grateful for any suggestions, remarks, ideas, etc. which might help me progress.

If you could leave the tape and the sheets of paper in my pigeon-hole or send them back to me (you have a stamped envelope with my address on it) as soon as possible, it would be great.

Thank you !

See you soon.

SOPHIE.

QUESTION TIME – Belfast**1^{er} passage :**

D.D.: On Question time from Belfast tonight Northern Ireland secretary Sir Patrick Mayhew ...(music).....leader of the SDLP John Hume(music).....Deputy leader of the Ulster unionists John Taylor(music).....and Labour's Northern Ireland spokeswoman Mo Mowlam(music).....(clapping, coughing etc.)..... Good evening Good evening Good evening good evening welcome welcome to Belfast and an audience ready with questions and arguments as always about the issues of the week and our first question comes from William Stephenson who's a retired civil servant Mr Stephenson

(paper rustling)W.S. Which would attract shorter odds from the bookmakers the prospect of a white Christmas or the resignation of the Chancellor of the exchequer ?

D.D. This is a kerfuffle today over whether the Chancellor had threatened to resign if n John Major changed his view on a single currency and a lot of talk about him having a row within the Tory party and telling them to telling the er (cough behind) chairman of the party to t get his get the scooters his kid's scooters off the lawn and all the rest of it Do you think there's a chance of the resignation of the Chancellor Mo Mowlam?

M.M. (sigh) Erm well I know in terms of shorter odds what I would like I'd like a white Christmas the resignation (little laugh) of Ken Clarke and an election so in terms of odds (laughter behind) I think my priorities would be quite clear I'm not they're so badly split and divided over it that there is a chance 'cause you just feel it in the House of Commons a lot the divisions are so fundamental and they're arguing so badly over Europe in the Tory Party that there is a pa chance that Erm there is a chance but I wouldn't put good odds on er him resigning to be straight with you I don't think that's a possibility at the moment.

D.D. *This is a kerfuffle today over whether the Chancellor had threatened to resign*

if n John Major changed his view on a single currency and a lot of talk about him having a row within the Tory party and telling them to

telling the er (cough behind) chairman of the party to t get his get the scooters his kid's scooters off the lawn and all the rest of it.

M.M. *(sigh) Erm well I know in terms of shorter odds what I would like I'd like a white Christmas the resignation (little laugh) of Ken Clarke and an election*

M.M. *:they're so badly split and divided over it*

there is a chance 'cause you just feel it in the House of Commons a lot

the divisions are so fundamental and they're arguing so badly over Europe in the Tory Party

that there is a pa chance that Erm there is a chance but I wouldn't put good odds on er him resigning to be straight with you I don't think that's a possibility at the moment.

2^{ème} passage :

D.D. Do you think he'd threaten to?

M.M. (hesitation) Well you see your guess is as good as mine on that I think that the lot of the information today suggests that he might have done and I think that the threats from either the er wing of the Tory Party are there and John Major is like a scared rabbit in the middle stuck as to to no movement on Europe so I think its possible that even if the words did not issue from his lips the equivalent might have done and don't forget there was also in the press today that about eighty five to a hundred Tories are going to have a different policy on Europe from the one of of their government so in that sense Erm its very unstable I don't think a resignation but I think it may mov mean an earlier election.

M.M. (hesitation) Well you see your guess is as good as mine on that

and I think that the threats from either the er wing of the Tory Party are there

and John Major is like a scared rabbit in the middle stuck as to to no movement on Europe

so I think its possible that even if the words did not issue from his lips

don't forget there was also in the press today that about eighty five to a hundred Tories are going to have a different policy on Europe

3^{ème} passage :

D.D. John Taylor?

J.T. Well We've had snow already in November which is unusual for us in Northern Ireland we'd snow even yesterday and er looking at it I think we're certainly going to have a better chance of snow on Christmas Day than the resignation of the Chancellor now having said that there's no doubt about it the Chancellor is an isolated man within the British Conservative party the views that he expresses does not represent the Conservative Party on the ground in Great Britain nor indeed in my opinion the Conservative Party in parliament but I think that if he did resign by Christmas that this government would collapse and we would have a general election and the only way in which this government can therefore remain in power is to maintain this rather uncomfortable coalition that exists within the cabinet in er which the Chancellor of the Exchequer in now very much a minority voice.

looking at it I think we're certainly going to have a better chance of snow on Christmas Day

than the resignation of the Chancellor

the views that he expresses does not represent the Conservative Party on the ground in Great Britain nor indeed in my opinion the Conservative Party in parliament

but I think that if he did resign by Christmas

and the only way in which this government can therefore remain in power

is to maintain this rather uncomfortable coalition that exists within the cabinet

4^{ème} passage :

D.D. Patrick Mayhew?

P.M. Fascinating insights into the cabinet which I I

(voice behind) Well you can tell us.....(giggles)

P.M. am a member of and I will I can tell you the odds are far stronger in favour of a white Christmas than having Ken Clarke resign and I'd say that without any reference to the fact that we've had our snow already there's a lot of creative accounting we hear about nowadays there's a great deal of creative journalism and this is creative

journalism there's a perfectly sensible policy which the Conservative cabinet the Conservative government and the Conservative party and the House of Commons are united behind to an overwhelming degree and that is that we're going to wait to see how this argument develops on a single currency we're not going to close the door on it now we're going to say well its going to affect us even if some of these people go in for it its going to affect us anyway so we're going to be in there helping to shape the argument and the discussion and when the time is right we'll have a look at it aside

I can tell you the odds are far stronger in favour of a white Christmas

there's a lot of creative accounting we hear about nowadays

there's a great deal of creative journalism and this is creative journalism

there's a perfectly sensible policy which the Conservative cabinet the Conservative government and the Conservative party and the House of Commons

and that is that we're going to wait to see how this argument develops on a single currency

we're not going to close the door on it now

we're going to say well its going to affect us even if some of these people go in for it its going to affect us anyway so we're going to be in there

helping to shape the argument and the discussion and when the time is right we'll have a look at it aside

5^{ème} passage :

D.D But yo yo you're one of the ones you're one of the few who is said to believe that in cabinet

P.M. Well I

D.D. I mean the reports by these er journalists who you so distrust ar eh the large number of people in cabinet don't think that and nearly and nearly change the Prime Minister forced him to change his hand.

P.M. News to me news to me news to me

D.D. Really news to you? Were you at cabinet last Thursday?

P.M. We have in Ken Clarke the best Chancellor for a generation he's got the thing that everybody's been looking for since the ending of the war low inflation rising employment falling unemployment and export lead boom everything that we've been looking for and that is the judgement of the man and led by the Prime Minister

D.D. And yet yet

P.M. The best Prime Minister and the best Chancellor and so lovely all all helps to sell the papers but it doesn't reflect the truth.

D.D. And and just before we leave it th th the three only out of twenty two members of the cabinet were said to be in favour of keeping the option open for a single currency

D.D. But yo yo you're one of the ones you're one of the few who is said to believe that in cabinet

D.D. I mean the reports by these er journalists who you so distrust

the large number of people in cabinet don't think that and nearly and nearly change the Prime Minister forced him to change his hand.

P.M. We have in Ken Clarke the best Chancellor for a generation

he's got the thing that everybody's been looking for since the ending of the war

low inflation rising employment falling unemployment and export lead boom

everything that we've been looking for and that is the judgement of the man and led by the Prime Minister

lovely all all helps to sell the papers but it doesn't reflect the truth.

th th the three only out of twenty two members of the cabinet were said to be in favour of keeping the option open for a single currency

6^{ème} passage :

?.?. The problem is the problem is let's face it you don't know whether you want a single currency or you don't want it you can't make you mind up and its time you did the nation wants to hear where you stand on this issue.

P.M. Well its a very interesting thing (clapping) It's a very interesting thing I I I wonder whether I wonder whether every single person here pretty representative of the nation pretty representative of the province I'm told anyway I wonder what everybody wants to say now we want to know absolutely firmly whether you're going in or whether you're going out never mind what the er circumstances are going to be in three or four years' time because if you say that I really don't think that you're representative of the Ulster people.

you don't know whether you want a single currency or you don't want it

you can't make you mind up and its time you did the nation wants to hear where you stand on this issue.

now we want to know absolutely firmly whether you're going in or whether you're going out never mind what the er circumstances are going to be in three or four years' time

7^{ème} passage :

Man: I'd like to ask Sir Patrick that if John Major did rule out joining a single European currency would he resign?

P.M.: If John Major ruled it out I would think that er he was no longer the wise Prime Minister which he has shown himself to be Now I don't th (sounds of general disagreement) I don't anticipate that happening so I'm not going to er take up the time of this programme by saying what I would do if something happened which I believe to be inconceivable

D.D.: If he ruled it out when? Patrick Mayhew

P.M.: Ooo well the suggestion is that if it was ruled out now what was asked but I was asked if it ruled it out now (others talking) would would I resign I think that that is so improbable that I'm not gonna even answer it he wouldn't do it he'd (hisses behind) be a totally different Prime Minister

Man: *I'd like to ask Sir Patrick that if John Major did rule out joining a single European currency would he resign?*

P.M.: *If John Major ruled it out I would think that er he was no longer the wise Prime Minister which he has shown himself to be*

Now I don't th (sounds of general disagreement) I don't anticipate that happening

so I'm not going to er take up the time of this programme by saying what I would do if something happened which I believe to be inconceivable

I think that that is so improbable that I'm not gonna even answer it he wouldn't do it he'd (hisses behind) be a totally different Prime Minister

8^{ème} passage :

D.D.: And Mo Mowlam do you think do you think if John Major had come out against

M.M.: Well if you'd just shut up and gave me a chance to answer (lots of audience laughter) I'd do it (huge applause) shall we shall we go again?

D.D.: Do you think do you think John Major was tempted to rule it out last Monday before he was smoked out on Tuesday

M.M.: I got a feeling I got a feeling that he was because if you looked at the poles that were in the Sun where people were saying (laughing behind) you doubting this well I'm just answering David's question I think that crossed his mind

P.M.: They all buy the Sun to look at the poles

?.?: Well listen leave

M.M.: Can I just answer

D.D.: What yes go on let Mo Mowlam answer (mutterings behind)

M.M.: I just wanted to answer because I keep getting nagged on my right here about what Labour's position is

?.?: Hmm right yeah ahahah

M.M.:... and our position in relation as it seems to be we are discussing the single currency here is our position is that a) there ought to be an referendum so people can actually debate discuss it and have their own say secondly in principle we believe in Europe we want a different relationship from the one that this lot have had because they've sat on the outside and John's absolutely right what we want to do is actually get in there and represent our country's interests and we've said that in relation to a single

currency its knee-jerk what you're saying that you're gonna go in or you're going to oppose it regardless or you're going to go in regardless we've said that in principle we'd like to be there but that we are going to wait and see next year what the state of the economy is whether the convergence criteria in relation to unemployment and the things that bother us are actually happening and that is will be the basis of our position

M.M.: I just wanted to answer because I keep getting nagged on my right here about what Labour's position is

M.M.:.. and our position in relation as it seems to be we are discussing the single currency here

our position is that a) there ought to be an referendum so people can actually debate discuss it and have their own say

secondly in principle we believe in Europe we want a different relationship from the one that this lot have had

what we want to do is actually get in there and represent our country's interests

and we've said that in relation to a single currency its knee-jerk what you're saying that you're gonna go in or you're going to oppose it regardless or you're going to go in regardless

we've said that in principle we'd like to be there

but that we are going to wait and see next year what the state of the economy is whether the convergence criteria in relation to unemployment and the things that bother us are actually happening

and that is will be the basis of our position

9^{ème} passage :

D.D. (...) Patrick Mayhew all there seem to be are talks about talks why?

P.M.: I don't think that's right the talks have been going on with nine out of ten parties elected for the purpose since the tenth of June a- and that took place oh not withstanding er Drumcree and er summer break and they're back there now now its slow progress but they are making progress and it is something which is very good news for the people of Northern Ireland who want to see their politicians talking but they want to see their politicians who are in there talking doing so on equal terms relying on their elected mandate and not upon weapons

D.D.: Heather Newell?

P.M.: *I don't think that's right*

and that took place oh not withstanding er Drumcree and er summer break and they're back there now

now it's slow progress but they are making progress

but they want to see their politicians who are in there talking doing so on equal terms relying on their elected mandate and not upon weapons

SONDAGES 2 et 3 :

Dear colleagues,

I'm still working on **emphasis** in English, and I would like to thank you for helping me. I need you to do another little poll. It is exactly the same principle as for the first one : I need native speakers to tell me what they find emphatic in the passages they hear. Just like in July, what you have to do is listen to a conversation (you will probably recognize colleagues) and then a radio recording and underline all the passages or words or syllables which you feel are **EMPHATIC**. I'm giving you the same definition : what I mean when I say emphatic is NOT NEUTRAL, with a special INVOLVEMENT on the part of the speaker. For example, ask yourselves if you would utter the same sentence in the same way in a neutral context. What I want you to tell me is what you think, what you feel is being made **prominent** in some way.

So again you have to listen to the tape (side A only) and fill in the paper that goes with it. It is shorter than the first one.

This time I would like you to distinguish at least two different degrees of emphasis (thanks to what you did in July) :

clear emphasis (C under passage or word underlined)

light emphasis (L under passage or word underlined)

You can add a third one (medium emphasis : M) if you feel the need.

So you first hear a short passage once and then I have selected sentences which you will hear twice and which are written in italics on the script. I would like you to underline on the text in italics the word(s) which you perceive as being emphatic or prominent. As for the first poll, I have left some space for you to make comments which might be useful to me. For instance hesitation or doubts or whatever might help me. Of course, not all the passages you are going to hear are emphatic, I mean you don't have to underline something in each sentence but you can if you think every sentence has something emphatic in it.

If you could leave the tape and the sheets of paper in my pigeon-hole as soon as possible, it would be great (I'm a bit in a hurry now...).

Thank you !

See you soon.

SOPHIE.

SONDAGE 2 : Conversation**Passage 1**

J : How did you find this camp site in Pertuis ?

V : Well because em when I as I knew I wanted to find somewhere in Pertuis and environs I em oh well that's not English I em I just went and you know and got in touch with the camp site and said could I come and stay and they said yes.

J : Did you do that last year as well ?

V : no last year I stayed with a friend er in her studio and er I think she did actually suggest that I stay there again this year but it it was not really in the area where I'm looking for something so it wasn't very convenient really. So er I'm doing it differently this year.

C : It's very difficult to find somewhere where you don't have to pay agencies' fees and everything else. It's like HHHH

V : And I mean just in one year, last year I was looking at this time you know this time last year I was looking and it's gone up the price of stuff has gone up

I just went and you know and got in touch with the camp site and said could I come and stay and they said yes.

J : Did you do that last year as well ?

no last year I stayed with a friend er in her studio

and er I think she did actually suggest that I stay there again this year

(but it it was) it's not really in the area where I'm looking for something so it wasn't very convenient really.

So er I'm doing it differently this year.

C : It's very difficult to find somewhere where you don't have to pay agencies' fees and everything else.

V : And I mean just in one year, last year I was looking at this time

you know this time last year I was looking and it's gone up

Passage 2

V : I found somewhere beautiful in Cadenet, absolutely beautiful at the top of Do you know Cadenet at all ?

J : Not really, I 've passed by.

V : well you know that ruined castle at the top

C : umm

V : well you know there's lots of winding little old streets going up there

C : yeah

V : it's right at the top so it looks out over the roofs of the village and you can even see la montagne Sainte Victoire

C : oh brilliant

V : and I would be able to pop up to the castle for a bit of exercise every evening and it's 2500 francs which I've come to consider as reasonably

J : That sounds very reasonable

V : well yes but it's tiny

V : I found somewhere beautiful in Cadenet, absolutely beautiful

V : well you know there's lots of winding little old streets going up there

V : it's right at the top

so it looks out over the roofs of the village

and I would be able to pop up to the castle for a bit of exercise every evening

J : That sounds very reasonable

V : well yes but it's tiny

Passage 3

V : Did you stay here over the summer ?

C : yeah yeah I was over here in the summer I went home in July for a couple of weeks and then I stayed here 'cause I was trying to decide on what topic to do for my mémoire for the er for the master's so I stayed over here but I didn't really take that much well I didn't really go out in the sun that much because it was too hot for me it was 40 degrees and I was just I went to the beach a couple of times but only in the morning

V : Why, oh because even that's too hot

C : it was just so hot and so full of people that I just thought No I'll stay at home in the countryside, I'm fine just have a little walk outside in the sun shine and then come back because it's just too much

V : So did you enjoy staying here or not ?

C : er I would have enjoyed it I think more if there'd been more people around like 'cause all the student friends I've made in the master's had gone home all the lecteurs and lectrices were off home visiting or weren't here yet so

J : I missed out on that heat wave

C : so I was a bit lonely

V : you like heat waves do you ?

J : well I'd like to try a heat wave

D : once in a while

C : It's good I think it's good if you're by a beach but if you're in town then it's not good like if you're in Aix and you have to travel like half an hour at least to go to the beach you spend your time in traffic jams with everybody else then it's not good . If you live by the beach then it's brilliant because you can just go down to the beach whenever you please but

I went home in July for a couple of weeks and then I stayed here

I didn't really go out in the sun that much because it was too hot for me it was 40 degrees

V : Why, oh because even that's too hot

C : it was just so hot and so full of people that I just thought No

J : I missed out on that heat wave

V : you like heat waves do you ?

J : well I'd like to try a heat wave

but if you're in town then it's not good like if you're in Aix and you have to travel like half an hour at least to go to the beach you spend your time in traffic jams with everybody else then it's not good .

Passage 4 :

C : I think it was just before I came back because I went back at the very end of June beginning of July and I rang my mum beforehand and said what clothes shall I bring and she said oh it's lovely it's gorgeous we have lovely weather so I took all my T shirts and my little skirts and my shorts and got over there and it was absolutely freezing

D : yeah the weather was really bad

C : I spent two weeks buying jumpers and

V : it's so difficult when you're somewhere hot to imagine what it's going to be like even if people tell you, isn't it ?

J : well when I got on the plane from Marseille going back to London and just bef I think half way through the flight they give you a progress report and the pilot said well there's a ten-degree difference between Marseille and and London so I thought oh

V : But even if they tell you you don't realise it's still ashock

J : It's still a shock

C : When I landed the pilot said oh so it's going to be about what 13 degrees which is quite good for this time of year I was like What ?

V : 13

so I took all my T shirts and my little skirts and my shorts and got over there and it was absolutely freezing

V : it's so difficult when you're somewhere hot to imagine what it's going to be like

I think half way through the flight they give you a progress report and the pilot said well there's a ten-degree difference between Marseille and and London so I thought oh

C : When I landed the pilot said oh so it's going to be about what 13 degrees which is quite good for this time of year

V : 13

Passage 5 :

C : 'cause his origins were in Yorkshire and obviously with mebeing at Bradford University then he came along and it was really nice he was a lovely lovely man honestly and he didn't bore anyone he just stood there making jokes the whole time [...]

J : So did you have anybody ?

V : Yeah I had the Duchess of Kent.

J/C : oh really

V : yea but actually I didn't go to my graduation ceremony ungrateful wretch that I am but I did for my PhD she gave me my PhD My mum felt that it was time that I went to one of these ceremonies. I think she wanted to buy a new dress

C : to buy a new dress (at the same time)

D : I think that's what it's for actually. Graduations are for parents because I really wasn't bothered about going but I know that my mum had been looking forward to it for years or whatever

J : oh ya I think that was about the only major celebration event in my life so I could invite parents and close friends and it was an excuse to buy a hat for them

C : exactly

D : I had a bit of a nightmare actually 'cause er before you know when you get the pack for that graduation thing where it says like measure your head and stuff so my mum measured my head but she must have done it wrong I don't know how she did it wrong but my mortar board thing was far too big for me so I could only walk about ...

C : you couldn't see

D : without moving my neck every time I moved my head the thing just came right down over my eyes like that so I had to I spent the whole time for the speeches sitting like this which is why the speeches dragged for me because of the 2 hours when my neck was killing me and then you had to go and then you had to nod to him don't you

J : Yeah

D : You shake his hand and nod and I was thinking oh my God I'm gonna nod and

C : If I nod it's gonna fall off

he came along and it was really nice he was a lovely lovely man honestly and he didn't bore anyone he just stood there making jokes the whole time

J : So did you have anybody ?

V : Yeah I had the Duchess of Kent.

V : yeah but actually I didn't go to my graduation ceremony ungrateful wretch that I am

but I did for my PhD she gave me my PhD

My mum felt that it was time that I went to one of these ceremonies.

I think she wanted to buy a new dress

J : oh ya I think that was about the only major celebration event in my life so I could invite parents and close friends and it was an excuse to buy a hat for them

D : I had a bit of a nightmare actually

so my mum measured my head but she must have done it wrong I don't know how she did it wrong but my mortar board thing was far too big for me

every time I moved my head the thing just came right down over my eyes like that

I spent the whole time for the speeches sitting like this which is why the speeches dragged for me 'cause of the 2 hours and my neck was killing me

and then you had to nod to him don't you

Passage 6 :

D : yeah but you know we did that with the flying the hat thing in the air and all of the photos we've got me and friends doing that all of the photos exactly the same with all of us with our hat our fingers all in the same position on one hand and our other hand is like that our fingers like that on the throwing hand and the other hand like that all of us exactly the same position it's really weird with our fingers

C : oh that's really good

D : it's really odd

V : in the same picture ?

D : yeah in the same picture we were all doing exactly the same things with our fingers and our other hand for some reason and we're all obviously looking up trying not to get hit on the head by...

V : It proves that higher education just turns out sort of imitations you know mould

C : yeah, everyone does the same thing we've all done the same degree now we're all doing exactly the same thing no it's good though university it gets you to spend so much time with different people I suppose you do take on other people's habits and stuff because you live with people for like four years or whatever so you do tend to sort of change a little bit I think at university

D : actually also which is particularly apt now that we're doing this emphasis thing is if you're living with someone you start to talk like that person talks as well and you emphasize your words in the way that they emphasize their words 'cause I was living with this guy and he had a particularly strange way of emphasizing his words and I realized after a while I was sort of doing the same thing

C : you were picking it up

D : sort of stressing the way he stressed words and stuff like that

D : yeah but you know we did that with the flying the hat thing in the air

all of the photos we've got me and friends doing that

all of us exactly the same position it's really weird with our fingers

no it's good though university

it gets you to spend so much time with different people

I suppose you do take on other people's habits and stuff because you live with people

for like four years or whatever

so you do tend to sort of change a little bit I think at university

if you're living with someone you start to talk like that person talks as well and you

emphasize your words in the way that they emphasize their words

sort of like stressing the way he stressed words and stuff like that

Passage 7

V : when you picked up French would you say that you were lucky enough never to have to be er exposed to people who were making mistakes and breaking up their sentences

C : no, not at all

J : what in French ?

V : I mean you're bound to aren't you ?

C : Yeah, you're bound to come across

[..]

V : I don't think it affects your grasp of the language

C : no unless I suppose if you spend an awful lot of time with that person you're gonna end up making horrendous mistakes if you don't know the language already you're just gonna be repeating their mistakes

V : I think people who learn a language just by picking it up you know as adults then they probably do pick up all sorts of mistakes but people like us who've studied the grammar and things I don't think

V : when you picked up French would you say that you were lucky enough never to have to be er exposed to people who were making mistakes and breaking up their sentences

V : I mean you're bound to aren't you ?

C : Yeah, you're bound to come across

V : I think people who learn a language just by picking it up you know as adults

then they probably do pick up all sorts of mistakes

SONDAGE 3 : BBC World Service

PASSAGE 1

BBC woman (Heather Paton): Earlier today I discussed the issue with two British Asian women. : Parro Tia (?), of gender watch, an organisation that looks at the problems of women and children in developing countries and the journalist and writer Yasmin Alibi Brown who was also involved in putting the report together. I put it to Yasmin that we had to make the distinction between arranged marriages and forced marriages.

Yasmin Alibi Brown: Well yes I think on the whole, arranged marriages the way they've developed in recent years is a consultation procedure between er parents and children and often they are remarkably more successful than love marriages because a lot of research goes into them. emm we're talking about forced marriages where any kind of coercion is used whether it's emotional blackmail or mothers threatening to commit suicide or actual abuse, physical violence and imprisonment. Any kind of coercion em which gets a person – and usually it is a young woman – into a position where she has to oblige to what the parents want is a forced marriage

BBC: How widespread a problem is it? I mean how many of these forced marriages take place in Britain every year?

YB: well if we It's a hidden problem. It's not even been discussed till this report came out properly as a journalist I would say my estimate would be something like 1500 to 2000 a year. What I do know is it's a hidden problem wi which communities here and on the subcontinent are keeping well hidden.

BBC: Do you think that the report is likely to achieve very much.

YB: Well I think the one thing it achieves just by existing is telling people this cannot be a hidden problem anymore. And actually it's quite hard hitting report in some ways you know where the Minister says quite rightly that cultural sensitivity is no excuse for moral blindness. Some things are wrong whoever does them. And that there are you know you can't make allowances or cultural religious allowances so in that way it's very hard hitting and it also liberates I think policemen social workers and teachers who have felt trapped that they can't interfere because they'll be accused for being racist. So I think in some ways it's an extremely significant report.

I put it to Yasmin that we had to make the distinction between arranged marriages and forced marriages.

and often they are remarkably more successful than love marriages because a lot of research goes into them.

we're talking about forced marriages where any kind of coercion is used whether it's emotional blackmail or mothers threatening to commit suicide or actual abuse, physical violence and imprisonment.

It's not even been discussed till this report came out properly

What I do know is it's a hidden problem wi which communities here and on the subcontinent are keeping well hidden.

Well I think the one thing it achieves just by existing is telling people this cannot be a hidden problem anymore.

Some things are wrong whoever does them.

It also liberates I think policemen social workers and teachers who have felt trapped that they can't interfere because they'll be accused for being racist.

So I think in some ways it's an extremely significant report.

Passage 9

BBC : Parro? what do you think of it ?

PS: I think there's an important principle here that er family and family relations are important and I think it is a private issue. Now

BBC : So you'd be against the idea of the authorities intervening

PS : Yes I'd be against the idea of government intervention or any third party intervention. I think we've got to recognize that precisely because it is a private issue one which I think can only be resolved within the family within friends relatives er any intervention er let it be social workers doctors er teachers er the government the police I think would cause more harm and er it would bring about distrust in the family it would weaken er all kinds of relationships

BBC : Yasmin has a horrified look

YB : I have a completely.. I mean would you say that about child abuse Would you say that the government has no business interfering in the privacy of a family with child because these are cases of child abuse they're girls of 13 and 14 we're talking about or would you say domestic violence is a private issue and you know because families are so precious we shouldn't interfere I think that's quite a shocking position to take

PS : You see I actually disagree I take the same position with child abuse and I think if you look at the last few years where child abuse first of all has been er defined er very very broadly everything and anything is now equated to child abuse and secondly we've actually seen quite a lot of er er cases where government has intervened it has backfired and the families are left in er you know quite a horrible state

any intervention er let it be social workers doctors er teachers er the government the police I think would cause more harm

Would you say that the government has no business interfering in the privacy of a family with child because these are cases of child abuse

or would you say domestic violence is a private issue

I think that's quite a shocking position to take

and secondly we've actually seen quite a lot of er er cases where government has intervened

it has backfired and the families are left in er you know quite a horrible state

Passage 10

BBC : Are you are you bothered about the cultural nature ? are you bo bothered about the the authorities of one culture intervening in another culture or are you just simply bothered by the idea of authorities intervening in families ?

PS : I'm bothered by the authority intervening in families

BBC : So if these were white families you you'd have exactly the same..

PS : I'd still say the same. Yes I would

BBC : You you actually you're you're talking from an interesting view point because you've had personally some of these sort of problems haven't you ?

PS : em Yes well em abouter twenty years ago when I decided to live with a a white person in this country my family almost disowned me er my father refused to speak to me and did not speak to me for five years er But I had friends I had er people to confide in and talk to and I must admit that er you know further down the line now my relationship with my family is very very strong and I think problems can be resolved. Er

BBC : if the authorities had intervened to help you when you perhaps needed help what would you have done ?

PS : Well I think you'd have to ask for help obviously I didn't go to the authorities but the authorities in as much as if they had intervened I think I would not today say that my relationship with my family could be as strong because I think intervention does create divisions pitches one against another the home office the foreign office I think order a leaflet to er in schools er explaining that you know when parents take their kids away to Pakistan and Bangladesh for long holidays there's aproblem watch out I think that makes every parent a subject for inquisition

YB : Yeah, and it's good and I'm very pleased it does because what we are saying here is that it's OK to force 14 year olds into marriage and have them legally raped I have cases where 14 and 15 year olds have been raped within these marriages until they were pregnant so they could be brought here and and I find it quite shocking that families these precious things called families are being you know treated like religions. All families are good when they are good and can be deeply oppressive and I think we should not never collude in the oppression of children and the violation of their their bodies and their rights in the name of this thing called family.

But I had friends I had people to confide in and talk to and I must admit that er you know further down the line now my relationship with my family is very very strong and I think problems can be resolved.

if the authorities had intervened to help you when you perhaps needed help what would you have done ?

obviously I didn't go to the authorities

but the authorities in as much as if they had intervened I think I would not today say that my relationship with my family could be as strong

because I think intervention does create divisions pictures one against another

I think that makes every parent a subject for inquisition

Yeah, and it's good and I'm very pleased it does because what we are saying here is that it's OK to force 14-year olds into marriage and have them legally raped

I have cases where 14 and 15 year olds have been raped within these marriages until they were pregnant so they could be brought here

and and I find it quite shocking that families these precious things called families are being you know treated like religions.

All families are good when they are good and can be deeply oppressive

and I think we should not never collude in the oppression of children and the violation of their their bodies and their rights in the name of this thing called family.

Annexe 2c - Script ‘TextGrid2Matrix’

```

#! perl

# TextGrid2Matrix.pl Perl droplet

# author: Daniel Hirst
# email: daniel.hirst@lpl.univ-aix.fr

# purpose: convert Praat TextGrid file to matrix

# input format:
# TextGrid with tiers "words" "prominence"
# output format:
# "$filename $entry $time $prosody"

# use prosody.pm in path "Macintosh HD:Perl:Modules"
use prosody;

#identify operating system and get directory separator $DS;
getds;
#get script name and time
($script = $0) =~ s/^\.*$DS//;
$now = localtime;

#set input and output directories
$indir = $ARGV[0];
$indirname = "TextGrid";
$outdirname = "Mtx";
-d $indir or die ("Sorry! $indir should be a directory\n");
($dir_path = $indir) =~ s/(\^.*){DS}$indirname$/1/ or die "Directory should be
[$indirname]\n";
opendir (INDIR, $indir) or die ("cannot open $indir $!");
@dir_list = readdir INDIR;
closedir INDIR;
$outdir = "$dir_path$DS$outdirname";
mkdir ($outdir, 0777);

FILE: for $file (@dir_list) {
    $infile = "$indir$DS$file";
    ($filename = $file) =~ s/\^.*//;
    next unless -T $infile; #skip files which are not text
    open (IN, $infile) or print "cannot open $file: $!", next FILE;
print "reading $file\n";
    $missing = 0;
    @line = ();
    while (<IN>) {

```

```

/^$/ and next;
/^;/ and next;
/^#/ and next;
/name = "(\\w+)/" and do {
    $tier = $1;
};
/xmin = "(\\d+\\.?\\d*)/" and do {
    $time = $1;
};
/text = "([a-zA-Z0-9\\.\\,\\/-=]*)\\s*/" and do {
    $text = $1 ? $1 : ".";
    if ($tier eq "words") {
        $words{$time} = $text;
    } elseif ($tier eq "prominence") {
        $prominence{$time} =
$text;
    } else {
        print "unknown tier $tier\\n";
    } #if
};
} #while (<IN>)
close (IN);

$outfile = "$outdir$DS$file";
open (OUT, ">$outfile");
print OUT "; created from $file by $script version $version on $now\\n";
print OUT "\\n";
for $t (sort keys %words) {
    print OUT "$filename\\t$words{$t}\\t$t\\t$prominence{$t}\\n";
} #for

close (OUT);
} #FILE: while

#version history
$version = "1.0";

#22feb2001 version 1.0 working

```

Annexe 2d - Script ‘AddPron2Matrix’

```

#!/ perl
# AddPron2Matrix.pl Perl droplet
# author: Daniel Hirst
# email: daniel.hirst@lpl.univ-aix.fr

# purpose: look up pronunciation in dictionary and add to matrix
# if prosody for word = "." then look for reduced form

# input format:
#      "$filename $entry $time $prosody
#
# output format:
#      "$filename $entry $time $prosody $transcription $info $nsylls

# use prosody.pm in path "Macintosh HD:logiciels sof:Perl:Modules"
use prosody;

#identify operating system and get directory separator $DS;
getds;
#get script name and time
($script = $0) =~ s/^\.*$DS//;
$now = localtime;

# read dictionaries:
$dic_path = "Macintosh HD:logiciels sof:Perl:reference";
for $dic_name ("oup-ald-s.dic", "EPG.dic", "transcription.dic", "reduced_forms.dic",
"garbage.dic") {
    $dic = "$dic_path$DS$dic_name";
    open (DIC, $dic) or die "Cannot open $dic_name $!";
    $nentries = 0;
    while (<DIC>) {
        chomp;
        ($entry, $pron, $info, $nsylls) = split /\t/; #entry can contain spaces
        $pron{$entry} = $pron;
        $info{$entry} = $info;
        $nsylls{$entry} = $nsylls;
        $nentries++;
    } #while
    close (DIC);
    print "$nentries entries read from $dic_name\n";
} #for $dic_name

#set input and output directories
$indir = $ARGV[0];
$indirname = "Mtx";

```

```

$outdirname = "Mtx+phon";
-d $indir or die ("Sorry! $indir should be a directory\n");
($dir_path = $indir) =~ s/(^.*){DS}$outdirname$/$1/ or die "Directory should be
[$outdirname]\n";
opendir (INDIR, $indir) or die ("cannot open $indir $!");
@dir_list = readdir INDIR;
closedir INDIR;
$outdir = "$dir_path$DS$outdirname";
mkdir ($outdir, 0777);

```

```

FILE: for $file (@dir_list) {
    $infile = "$indir$DS$file";
    next unless -T $infile; #skip files which are not text
    open (IN, $infile) or print "cannot open $file: $!", next FILE;
print "treating $file\n";
    $missing = 0;
    @line = ();
    while (<IN>) {
        /^$/ and do {push @line, $_; next;};
        /^;/ and do {push @line, $_; next;};
        /^#/ and do {push @line, $_; next;};
        chomp;
        ($filename, $entry, $time, $prosody) = split /\t/; #entries can contain
spaces
        if ($entry2 = check ($entry, $prosody, \%pron)) {
            $entry = $entry2;
            $transcription = $pron{$entry};
            $info = $info{$entry};
            $nsylls = $nsylls{$entry};
        } else { #entry not in dictionary
            @words = ();
            $transcription = "";
            $info = "";
            $nsylls = 0;
            splitword($entry, \@words);
            $entry = "";
            for $word (@words) {
                if ($word2 = check ($word, $prosody, \%pron)) {
                    $entry .= "$word2 ";
                    $transcription .= $pron{$word2};
                    $info .= $info{$word2};
                    $nsylls += $nsylls{$word2};
                } else { #word not in dictionary
                    $entry .= $word;
                    $transcription = "***";
                    $info = "***";
                    $nsylls = 0;
                    $missing++;
                } #if
            } #for $word
            $entry =~ s/ $//;

```

```

        } #if
            push @line,
"$filename\t$entry\t$time\t$prosody\t$transcription\t$info\t$nsylls\n";
    } #while (<IN>)
close (IN);

$outfile = "$outdir$DS$file";
open (OUT, ">$outfile");
print OUT "; pronunciation etc added to $file by $script version $version on $now\n";
if ($missing) {
    $s = ($missing == 1) ? "" : "s";
    print "**** file $file contained $missing word$s not in the
dictionaries****\n";
}
foreach (@line) {print OUT};
} #FILE: while

sub check {
    my ($entry, $prosody, $pronref) = @_ ;
    my ($reduced, $lowered, $both);
    $reduced = "_$entry";
    $lowered = lc($entry);
    $both = lc($reduced);
    if ($prosody ne '.') { #stressed word
        if (defined($$pronref{$entry})) {
            return $entry;
        } elsif (defined($$pronref{$lowered})) {
            return $lowered;
        } else {
            return 0;
        }
    } elsif (defined($$pronref{$reduced})) { #unstressed word
#print "$entry      reduced\n";
        return $reduced;
    } elsif (defined($$pronref{$both})) {
        return $both;
#print "$entry      reduced and lowered\n";
    } elsif (defined($$pronref{$entry})) {
        return $entry;
#print "$entry      lowered\n";
    } elsif (defined($$pronref{$lowered})) {
        return $lowered;
    } else {
#print "$entry      not in dictionary\n";
        return 0;
    }
} #sub check
#version history
$version = "1.1";
#14feb2001 version 1.1 check entry in dictionary before splitting word
#10feb2001 version 1.0 working

```

Annexe 3a - Script ‘measure F0’

```

#Praat script
#name: measure F0 for labels
#calculates for each label
#   minimum F0 (Hz)
#   time minimum F0 (s)
#   maximum F0 (Hz)
#   time maximum F0 (s)
#   mean F0 (semitones)
#   standard deviation (semitones)
#   mean absolute slope (semitones)
#   mean absolute slope without octave jumps
#author: Daniel Hirst
#email: daniel.hirst@lpl.univ-aix.fr
#date: 4 May 2001
#latest revision: 4 May 2001
# requires TextGrid containing tiers UI, UT and word as tiers 1 to 3
# requires Sound with same name as TextGrid
# sound must be selected
# labels for Pause on each tier must begin with "P" (they will be skipped)
# labels corresponding to a duration less than 40 ms will also be skipped
name$ = selected$("Sound")
Edit
clearinfo
print F0 information for file 'name$'
printline
print MinF0 TMinF0 MaxF0 TMaxF0 Mean SD MAS MASWOJ
printline
for tier from 1 to 3
print Tier 'tier'
printline
select TextGrid 'name$'
nint = Get number of intervals... 'tier'
for i from 1 to nint
    select TextGrid 'name$'
    label$ = Get label of interval... 'tier' 'i'
    start = Get starting point... 'tier' 'i'
    stop = Get end point... 'tier' 'i'
    duration = stop - start
    if left$(label$,1) <> "P" and duration > 0.04
        editor Sound 'name$'
            Move B to... start
            Move E to... stop
            Extract selection (preserve times)
        endeditor
    To Pitch... 0.01 75 300

```

```

Rename... raw
Smooth... 10
Rename... smooth
minf0 = Get minimum... 0 0 Hertz Parabolic
tminf0 = Get time of minimum... 0 0 Hertz Parabolic
maxf0 = Get maximum... 0 0 Hertz Parabolic
tmaxf0 = Get time of maximum... 0 0 Hertz Parabolic
mean = Get mean... 0 0 Semitones
sd = Get standard deviation... 0 0 Semitones
mas = Get mean absolute slope... Semitones
maswoj = Get slope without octave jumps
print 'label$' 'minf0' 'tminf0' 'maxf0' 'tmaxf0' 'mean' 'sd' 'mas' 'maswoj'
printline
select Sound untitled
plus Pitch raw
plus Pitch smooth
Remove

endif
endfor #i
endfor #tier

```


Annexe 3b - 'measure intensity'

```

#Praat script
#name: measure intensity for labels
#calculates for each label
#       minimum int (dB)
#       time minimum int (s)
#       maximum int (dB)
#       time maximum int (s)
#       mean int (dB)
#       standard deviation (dB)
#author: Sophie Dujardin
#email: dujardin@aixup.univ-aix.fr
#date: 30 May 2001
#latest revision: 1 June 2001
# requires TextGrid containing tiers UI, UT as tiers 2 to 3
# requires Sound with same name as TextGrid
# sound must be selected
# labels for Pause on each tier must begin with "P" (they will be skipped)
# labels corresponding to a duration less than 64 ms will also be skipped
name$ = selected$("Sound")
Edit
clearinfo
print intensity information for file 'name$'
printline
print MinInt TMinInt MaxInt TMaxInt MeanInt SDInt
printline
for tier from 1 to 3
print Tier 'tier'
printline
select TextGrid 'name$'
nint = Get number of intervals... 'tier'
for i from 1 to nint
    select TextGrid 'name$'
    label$ = Get label of interval... 'tier' 'i'
    start = Get starting point... 'tier' 'i'
    stop = Get end point... 'tier' 'i'
    duration = stop - start
    if left$(label$,1) <> "P" and duration > 0.064
        editor Sound 'name$'
            Move B to... start
            Move E to... stop
            Extract selection (preserve times)
        endeditor
        To Intensity... 100 0.01
        Rename... raw
        minInt = Get minimum... 0 0 Parabolic
    
```

```

    tminInt = Get time of minimum... 0 0 Parabolic
    maxInt = Get maximum... 0 0 Parabolic
    tmaxInt = Get time of maximum... 0 0 Parabolic
    mean = Get mean... 0 0
    sdInt = Get standard deviation... 0 0
    print 'label$' 'minInt' 'tminInt' 'maxInt' 'tmaxInt' 'mean' 'sdInt'
    printline
    select Sound untitled
    plus Intensity raw
    Remove
endif
endfor #i
endfor #tier

```

Annexe 3c - Script 'mtostat'

```
# Matrix to .stat for Excel analysis
# Author : Cyril Auran
# E-mail : cyril.auran@free.fr

#identify operating system
for ($^O) {
    /MacOS/ and do {$DS = ":" , last;};
    /Unix/ and do { $DS = "\"" ; last;};
    /MSWin32/ and do { $DS = "\" " ; last;};
    die "You need to define the directory separator for $^O";
}

($droplet_name = $0) =~ s/^.*$DS//;

#read duration stats
open (STA, "Macintosh HD:logiciels sof:Perl:reference:dur-s.stats") or die ("Cannot
open stat file\n");
$a=0;
$z=0;
while (<STA>) {
    chomp;
    unless (/^;/) {
        ($pho[$a], $z, $meantime[$a], $standarddev[$a])=split(/\t/);
        $mean{$pho[$a]}=$meantime[$a];
        $sd{$pho[$a]}=$standarddev[$a];
        if (length($pho[$a])>1) {
            splice(@dipht, $z, 0, $pho[$a]);
            $z++;
        }
        $a++;
    }
}
close (STA);

# input directory+files and output directory+files
$indir = "Macintosh HD:Thèse Sof:Mtx+phon";
-d $indir or die ("Sorry! $indir should be a directory\n");
opendir (INDIR, $indir) or die ("cannot open $indir!");
print "\nthinking about it ...\n";
@dir_list = readdir INDIR;
closedir INDIR;
$outdir = "Macintosh HD:Thèse Sof:Stat";
mkdir ($outdir, 0777);

# error logbook
```

```

$logb="errors.txt";
open (ERR, ">$outdir$DS$logb");
print "error logbook opened\n";

#processing each file in turn
FILE: while ($file = pop(@dir_list)) {
    $infile = "$indir$DS$file";
    open (IN, $infile) or print "cannot open $file: $!", next FILE;
    ($outfile = $file) =~ s/{3}$ /stat/;
    open (OUT, ">$outdir$DS$outfile");

    #initiating index
    $b=0;

    # processing files
    while (<IN>){
        chomp;
        unless (/^;/) {
            ($file[$b], $word[$b], $time[$b], $pros[$b], $pron[$b], $pos[$b],
            $nsyl[$b])=split (/t/);
            $b++;
        }
    }
    print "Processing file $file\n";
    print OUT "$file analysed by mtostat.pl on ";
    print OUT scalar(localtime);
    print OUT "\n";
    print OUT "WORD/ACTUAL DURATION/STATISTICAL MEAN
DURATION/LENGTHENING (ms/%) \n";

    $v='[iIeO{AQoUuV3\@]';

    for ($c=0; $c<($b-1); $c++) {
        $pron[$c]=~s/^'/;
        $pron[$c]=~s/^./;
        if ($pron[$c]=~s/^*/) {
            $pron[$c]=~s/^*$/;
            if ($pron[$c]=~/[sz]s$/) {
                $pron[$c]=~s/s$/Iz/;
            }
            if ($pron[$c]=~/[ptkfst]s$/) {
                $pron[$c]=~s/s$/s/;
            }
            else {
                $pron[$c]=~s/s$/z/;
            }
            if ($pron[$c+1]=~/^[\\,]*$v+/) {
                $pron[$c]=~s/^DIS$/Di:/;
            }
            else {
                $pron[$c]=~s/^DIS$/D@/;
            }
        }
    }
}

```

```

        $pron[$c]=~s/r$//;
    }
}
if ($pron[$c]=~/.*\*.*+/) {
    $pron[$c]=~s/\*//;
}

$dur{$word[$c]}=1000*($time[$c+1]-$time[$c]);
&chunk($pron[$c]);

print OUT "\n$word[$c]\t";
printf OUT ("%10.3f", "$dur{$word[$c]}");
print OUT "\t";
printf OUT ("%10.3f", "$sig");
print OUT "\t";

$dif=$dur{$word[$c]}-$sig;
if ($sig!=0) {
    $len=$dif*100/$sig;
    printf OUT ("%10.3f", "$dif");
    print OUT "\t";
    printf OUT ("%10.2f", "$len");
}
else {
    print OUT "UNKNOWN DUR";
    print ERR "$file\t $word[$c]\t $dur{$word[$c]}\t UNKNOWN
DUR\n";
}

$dif=0;
$len=0;
%dur=();
@phon=();

}

close (IN);
close (OUT);
}

close (ERR);
print "error logbook closed";

#analysing the pron
sub chunk {
    my ($phon)=@_;
    @phon=split(//, $phon);

    my $sigmean;

    for ($i=0 ; $i<=$#phon; $i++) {
        for $j (0..$#dipht) {

```

```

        if ("Sphon[$i]Sphon[$i+1]" =~ /($dipht[$j])/) {
            splice(@phon, $i, 2, join(",$phon[$i],$phon[$i+1]"));
            last;
        }
    }
}

foreach $phon (@phon) {
    $phon=~s/5/l/;
}

for ($i=0; $i<=$#phon; $i++) {
    $sigmean += $mean{$phon[$i]};
}

$sig=$sigmean;
}

# Version history :
# 13/03/01 : working
# 14/03/01 : file name added
# 15/03/01 : absolute lengthening added
# 08/07/01 : changes for SD -

```

Annexe 3d - Intensité : pré analyses

Mesures faites manuellement avec PRAAT : du son → to « intensity » puis « down to intensity tier ».

Question Time :

Locuteur DD : 7 passages analysés :

P1.1S01 :

- fort : 40dB

+ fort: 75,8 dB

Cons sourdes : 40-45

Voy réduites : 63-64

Voy inacc : 70-72

Voy acc : 70-74

2 mots emphatiques : pas les plus forts. Se comportent comme des mots normaux.

P1.1S02 :

- fort : 42dB

+ fort: 73,8 dB.

Ces deux valeurs sont très constantes

Cons sourdes : 40

Voy réduites : 59

Voy inacc : 68

Voy acc : 73

1 mot emphatique : pas plus fort. Se comporte comme les mots non emphatiques.

P1.1S03 :

- fort : 38,3 dB

+ fort: 75 dB

Cons sourdes : 40

Voy réduites : 58

Voy inacc : 68

Voy acc : 73

3 mots emphatiques : pas plus forts. Se comportent comme des mots non emphatiques (scooters (2) plutôt moins fort).

P1.5S01 :

- fort : 43dB

+ fort: 76,75 dB. **Car une personne parle en même temps au début donc DD parle plus fort pour avoir la parole.**

Cons sourdes : 45

Voy réduites : 61 ?

Voy inacc : 70-72

Voy acc : 74-75

1 mot emphatique : un peu moins fort que les mots les plus forts (75,86dB) car la personne qui parlait en même temps vient de se taire.

P1.5S02 :

- fort : 37 dB

+ fort: 77,75 dB **Car une personne parle en même temps au début donc DD parle plus fort pour avoir la parole.**

Cons sourdes : 42

Voy réduites :

Voy inacc : 70

Voy acc : 73-75

1 mot emphatique : plus fort « reports » (76,23 / 77,75 dB)

P1.5S03 :

- fort : 40,5 dB

+ fort: 75,9 dB

Cons sourdes : 41

Voy réduites : 60 ?

Voy inacc : 70

Voy acc : 73

2 mots emphatiques : plus forts.

P1.5S09 :

- fort : 39 dB

+ fort: 76,4 dB **Début fort car un autre parle en même temps.**

Cons sourdes : 42-46

Voy réduites : 62

Voy inacc : 69-71

Voy acc : 73

1 mot emphatique : pas plus fort (74,34dB).

Locuteur MM (19 passages)

P1.1S04 :

- fort : 39 dB

+ fort: 77,5 dB (mais rire) sinon 75,3 (premier mot)

Cons sourdes : 42

Voy réduites : 37

Voy inacc : 70

Voy acc : 72

1 mot emphatique : plutôt moins fort : « would » : 69,7dB

P1.1S05 :

- fort : 42,75dB

+ fort: 74,4 dB

Cons sourdes : 45

Voy réduites : 63 (une seule)

Voy inacc : 68

Voy acc : 73

1 mot emphatique : 73,59dB normal.

P1.1S06 :

- fort : 38 dB

+ fort: 76,4 dB (premier mot)
 Cons sourdes : 40
 Voy réduites : 62
 Voy inacc : 70
 Voy acc : 72

P1.1S07 :

- fort : 39,5 dB
 + fort: 74,3 dB (premier mot et mot emph)
 Cons sourdes : 40
 Voy réduites : 37
 Voy inacc : 69-71
 Voy acc : 72-73
 2 mots emphatiques : « so » :74,8 dB (aussi fort que « divisions » (head)) « badly » :73,8dB normal.

P1.1S08 :

- fort : 39 dB
 + fort: 78,2 dB (premier mot)
 Cons sourdes : 40
 Voy réduites : 66
 Voy inacc : 70
 Voy acc : 72,5-73,5
 3 mots emphatiques : normaux

P1.2S01 :

- fort : 39 dB
 + fort: 73 dB (mot emph)
 Cons sourdes : 41
 Voy réduites : 65
 Voy inacc : que des monosyll
 Voy acc : 71
 1 mot emphatique : plus fort (73dB)
 Deuxième mot le plus fort pas donné emph du tout

P1.2S02 :

- fort : 39 dB
 + fort: 78 dB (wing emph : 22,22 mais hésitation juste avant)
 Cons sourdes : 40
 Voy réduites : 70
 Voy inacc : 70-71
 Voy acc : 73-74
 1 mot emphatique : 76,8 dB (un peu plus fort que les autres mais moins fort que « wing »)

P1.2S03 :

- fort : 39 dB
 + fort: 75,5 dB (mot emph)
 Cons sourdes : 40
 Voy réduites : 65
 Voy inacc : 70
 Voy acc : 74,5
 2 mots emphatiques : pas + forts (« stuck » vers 74 dB et « movement » plutôt moins fort).

P1.2S04 :

- fort : 41 dB
 + fort: 77,9 dB (« possible »)
 Cons sourdes : 41

Voy réduites : 66
 Voy inacc : 71
 Voy acc : 74

P1.2S05 :

- fort : 38,2 dB
 + fort: 76,5 dB (« gonna »)
 Cons sourdes : 38-39
 Voy réduites : 67
 Voy inacc : 70
 Voy acc : 73-74
 1 mot emphatique : pas + fort (different : 73,74 dB normal)

P1.8S01 :

- fort : 42,15 dB
 + fort: 74,3 dB (head of the into unit : « just »)
 Cons sourdes : 45
 Voy réduites : 64
 Voy inacc : 67
 Voy acc : 72
 1 mot emphatique : deuxième + fort (74,01 dB)

P1.8S02 :

- fort : 42 dB
 + fort: 74,5 dB (head of the into unit: « our »)
 Cons sourdes : 45
 Voy réduites : 65
 Voy inacc : 67
 Voy acc : 71-72

P1.8S03 :

- fort : 40 dB
 + fort: 76,35 dB (mot emph)
 Cons sourdes : 40-44
 Voy réduites : 60-66
 Voy inacc : 67
 Voy acc : 69,5-71,5
 2 mots très emphatiques : 74,93 & 76,35 dB.
 Les autres un peu emph sont dans les plus forts.

P1.8S04 :

- fort : 40 dB
 + fort: 77 dB (mot emph)
 Cons sourdes : 40
 Voy réduites : 66
 Voy inacc : 70
 Voy acc : 73
 mots emphatiques dans les + forts (sauf « different » plutôt moins fort)

P1.8S05 :

- fort : 40 dB
 + fort: 74,6 dB (premier mot)
 Cons sourdes : 40
 Voy réduites : 62-66
 Voy inacc : 69-70
 Voy acc : 73
 mot emphatique deuxième plus fort (« in » : 74 dB)

P1.8S06 :

- fort : 41,2 dB
 + fort: 76,75 dB (mot emph)
 Cons sourdes : 44-46
 Voy réduites : 64-66
 Voy inacc : 67-68
 Voy acc : 71
 mots emphatiques plus forts (vers 73 dB plutôt que 71)

P1.8S07 :

- fort : 43,3 dB
 + fort: 77 dB (mot emph : « principe »)
 Cons sourdes : 44,5
 Voy réduites : 63 (une seule)
 Voy inacc : 69
 Voy acc : 72
 Deuxième mot emph pas plus fort (vers la fin)

P1.8S08 :

- fort : 39,5 dB
 + fort: 76,5 dB (mot emph : what)
 Cons sourdes : 41-42,5
 Voy réduites : 60-66
 Voy inacc : 68,5
 Voy acc : 70
 mots emphatiques dans les plus forts (vers 73 dB)

P1.8S09 :

- fort : 37 dB
 + fort: 75,43 dB (mot emph)
 Cons sourdes : 44
 Voy réduites : 60 (une seule)
 Voy inacc : 61-62
 Voy acc : 71
 mots un peu emphatiques dans les plus forts

Locuteur JT (9 passages)P1.3S01 :

- fort : 39,5 dB
 + fort: 73,71 dB (mot emph)
 Cons sourdes : 42
 Voy réduites : 63-67 ?
 Voy inacc : 68
 Voy acc : 71
 2 mots emphatiques : les 2 plus forts

P1.3S02 :

- fort : 39,3 dB
 + fort: 72,25 dB (head sur acc sec)
 Cons sourdes : 39-43
 Voy réduites : 67
 Voy inacc : 70
 Voy acc : 71
 1 mot emphatique : 71 dB

P1.3S03 : non emph

- fort : 38 dB
 + fort: 69,3 dB (head sur acc sec)
 Cons sourdes : 38-46
 Voy réduites :
 Voy inacc : 61

Voy acc : 68

Moins fort que les autres passages (car pas du tout emph ?)

P1.3S04 :

- fort : 35,7 dB
 + fort: 74 dB (head)
 Cons sourdes : 37-38
 Voy réduites : 67
 Voy inacc : 68,5
 Voy acc : 71
 1 mot emphatique : 71,72 dB (un tout petit peu plus fort)

P1.3S05 :

- fort : 39,3 dB
 + fort: 75 dB (head et mot emph : did)
 Cons sourdes : 39-43
 Voy réduites : 67
 Voy inacc : 70
 Voy acc : 70- 71
 2 mots emphatiques : « did » le plus fort et « Christmas » pas fort : normal

P1.3S06 : non emph

- fort : 38,3 dB
 + fort: 70,8 dB (head et mot emph : did)
 Cons sourdes : 38,3
 Voy réduites :
 Voy inacc : 63,2
 Voy acc : 68-70,8

Pas très fort car non emph ?

P1.3S07 :

- fort : 40 dB
 + fort: 72,65 dB (mot emph)
 Cons sourdes : 40-43
 Voy réduites :
 Voy inacc : 67-69
 Voy acc : 71
 mot emphatique le plus fort

P1.3S08 :

- fort : 37,8 dB
 + fort: 73,7 dB (head et mot emph : did)
 Cons sourdes : 40-42
 Voy réduites : 65
 Voy inacc : 66
 Voy acc : 70- 71
 mot emphatique le plus fort et « within » deuxième plus fort

P1.3S09 : non emph (assez monotone)

- fort : 39,9 dB
 + fort: 71 dB (head et mot emph : « did »)
 Cons sourdes : 40-42
 Voy réduites : 59
 Voy inacc : 60
 Voy acc : 66-68

Pas très fort car non emph ?

NOTE : Les articles réduits (« a » et « the ») sont plus forts en intensité que toutes les autres voyelles réduites (comme « of ») chez les trois locuteurs.

CONCLU : passages non emph moins forts que les autres en moy.

Locuteur VB (17 passages)

P2.1S01 : non emph (déclin de l'intensité sur 2^{ème} partie de la phrase. Commence fort et baisse en intensité. Uniquement des monosyll.)
 - fort : 34,2 dB(dernier mot)
 + fort: 73,5 dB (head) 65,5 sur la deuxième partie
 Cons sourdes : 35-39
 Voy réduites : 62
 Voy inacc :
 Voy acc : 60-65

P2.1S03 : déclin de l'intensité
 - fort : 34,2 dB(dernier mot)
 + fort: 74,8 dB (no et mot emph)
 Cons sourdes : 36,5
 Voy réduites : 64,5
 Voy inacc : 68 (une seule)
 Voy acc : 70-73

Mot emph le plus haut avec no

P2.1S04 :
 - fort : 38 dB(dernier mot)
 + fort: 71,9 dB (no et mot emph)
 Cons sourdes : 40-45
 Voy réduites : 59
 Voy inacc : 62-66
 Voy acc : 69,8

Mot emph pas le plus haut mais haut comme les V acc.

P2.1S05 : déclin de l'intensité sur deuxième partie. non emph
 - fort : 36,6 dB
 + fort: 74,15 dB (head et area)
 Cons sourdes : 40-42
 Voy réduites : 56
 Voy inacc : 64,5/65,5-67
 Voy acc : 65,5-70/71

P2.1S06 : Non emph. Pas très fort. Remonte un peu vers la fin
 - fort : 39 dB
 + fort: 65 dB (head et « area »)
 Cons sourdes : 40-44
 Voy réduites : 59
 Voy inacc : 62-63
 Voy acc : 60-65
 Semble un peu tout pareil car pas du tout emph.

P2.1S08 : déclin de l'intensité sur deuxième partie

- fort : 37,3 dB(dernier mot)
 + fort: 76,7 dB (head)
 Cons sourdes : 40-42
 Voy réduites : 60-66
 Voy inacc : 65 (une seule)
 Voy acc : 73

2 Mots emph pas les plus hauts mais hauts comme les V acc.

P2.1S09 :
 - fort : 33 dB(dernier mot)
 + fort: 71,5 dB (n de gone)
 Cons sourdes : 33-36
 Voy réduites : 64
 Voy inacc : 64 (une seule)
 Voy acc : 64-67

1 mot emph : « up » (vient après « gone » et c'est le nucleus qui est le plus fort !)

P2.2S01 : tendance au déclin de l'intensité.
 - fort : 35,2 dB
 + fort: 78,9 dB (head) mais bruit plutôt 73
 Cons sourdes : 44
 Voy réduites :
 Voy inacc : 60
 Voy acc : 65-73

1 mot emph à 100% : pas plus fort. Tendance quand-même à plus d'intensité sur mots en relief car plusieurs dans cette phrase mais pas en corrélation avec degré d'emph.

P2.2S02 : tendance au déclin de l'intensité. Non emph
 - fort : 37 dB
 + fort: 73 dB (head)
 Cons sourdes : 40-43
 Voy réduites : 63
 Voy inacc : 69 ?
 Voy acc : 69-71

P2.2S03 :
 - fort : 35,5 dB
 + fort: 74,8 dB (head qui est aussi le mot emph)
 Cons sourdes : 35,5-45
 Voy réduites : 67,5
 Voy inacc
 Voy acc : 69-74,8 (que 2)

P2.2S04 : déclin de l'intensité sur dernier mot. Non emph
 - fort : 37 dB
 + fort: 73,74 dB
 Cons sourdes : 40-43
 Voy réduites : 62-70
 Voy inacc :
 Voy acc : 71

P2.2S05 : Non emph

- fort : 33,7 dB
 + fort: 74 dB (head)
 Cons sourdes : 34-38
 Voy réduites : 64
 Voy inacc : 64
 Voy acc : 67-70

P2.2S07 : 2 mots acc dont 1 emph : pareil
 - fort : 34,4dB
 + fort: 72,68 dB (head)
 Cons sourdes : 34
 Voy réduites : 64
 Voy inacc : 67
 Voy acc : 72,68

P2.5S03 : Emph : déclin :
 - fort : 33,9 dB
 + fort: 75,9 dB (head)
 Cons sourdes : 34-35
 Voy réduites : 59,5
 Voy inacc : 63
 Voy acc : 72/73
 Mot emph le plus bas (64 dB). L'emph se fait grâce à la baisse d'intensité dans ce cas précis. Peut-être parce qu'au début, elle monte pour se faire entendre car on parle en même temps qu'elle.

P2.2S04 : Déclin
 - fort : 41 dB
 + fort: 77,9 dB (ungrateful)
 Cons sourdes : 41-42

Voy réduites : 60
 Voy inacc : 67
 Voy acc : 67-74
 Mot emph haut mais un peu moins que « ungrateful »

P2.2S05 : Déclin 2^{ème} partie
 - fort : 40 dB
 + fort: 76,7 dB (head et mot emph : le même)
 Cons sourdes : 40-49
 Voy réduites :
 Voy inacc :
 Voy acc : 71-60
 Mot emph haut mais un peu moins que ungrateful

P2.2S06 : Non emph Déclin
 - fort : 37,9 dB
 + fort: 70,6 dB (head)
 Cons sourdes : 44,5
 Voy réduites :
 Voy inacc : 63
 Voy acc : 66

CONCLU VB_i :
Tendance au déclin de l'intensité sur la phrase entière. Le head est svt le plus intensif, pas le mot emph, qui est plutôt comme des syll acc normales. Un cas où le mot emph est bcp moins intensif que le reste (procédé pour emph ?) passsages non emph moins forts que les autres en moy

Annexe 3e - Nombre de pauses

Nombre de pauses par loc :

locuteur	nbre pauses
DD	8
MM	19
JT	1
PM	20
VB	6
JC	4
CB	10
DL	6
HP	3
YB	38
PS	16

Distribution des pauses en pourcentage :

Locu- teur	% pauses int	% pauses hes	% pauses emph
DD	37,5	25	50
MM	78,947	26,31	42,105
PM	55	20	50
VB	50	50	33,333
CB	80	10	20
DL	50	0	66,666
YB	36,842	34,210	39,473
PS	43,75	50	25
Total	54,004	26,94	40,822

Pourcentage des pauses int par loc :

locuteur	% pauses int
DD	37,5
MM	78,947
PM	55
VB	50
CB	80
DL	50
YB	36,842
PS	43,75

Pourcentage des pauses hes par loc :

locuteur	% pauses hes
DD	25
MM	26,315
PM	20
VB	50
CB	10
DL	0
YB	34,21
PS	50

Pourcentage des pauses emph par loc :

Locuteur	% pauses emph
DD	50
MM	42,105
PM	50
VB	33,33
CB	20
DL	66,66
YB	39,473
PS	25

Annexe 3f - Durée des pauses

Cf. CD ROM : *Pauses* pour les détails des durées de chaque pause.

Durées moyennes des trois types de pauses (msec) :

pauses int	348,046929
pauses hes	327,799239
pauses emph	335,705729

Durées moyennes des pauses int, hes et emph par locuteur :

Locuteur	pauses int	pauses hes	pauses emph
DD	285,133333	337,5	313
MM	307,533333	484,9	241,8375
FM	364,590909	197,025	289,51
VB	490,6	846,033333	669,5
CB	367,175	189	260,5
DL	345,9		320,825
YB	296,457143	313,723077	395,473333
PS	326,985714	254,2125	195

Annexe 3g - Pauses emphatiques, avant ou après

Pourcentage des pauses emphatiques avant et après par locuteur :

Locuteur	% pauses emph avant	% pauses emph après
DD	50	50
MM	25	75
PM	50	60
VB	50	50
JC	100	0
CB	50	50
DL	75	25
HP	50	50
YB	60	53,33
PS	100	0

Annexe 4a - Liste des abréviations pour les variables utilisées avec CRUISE

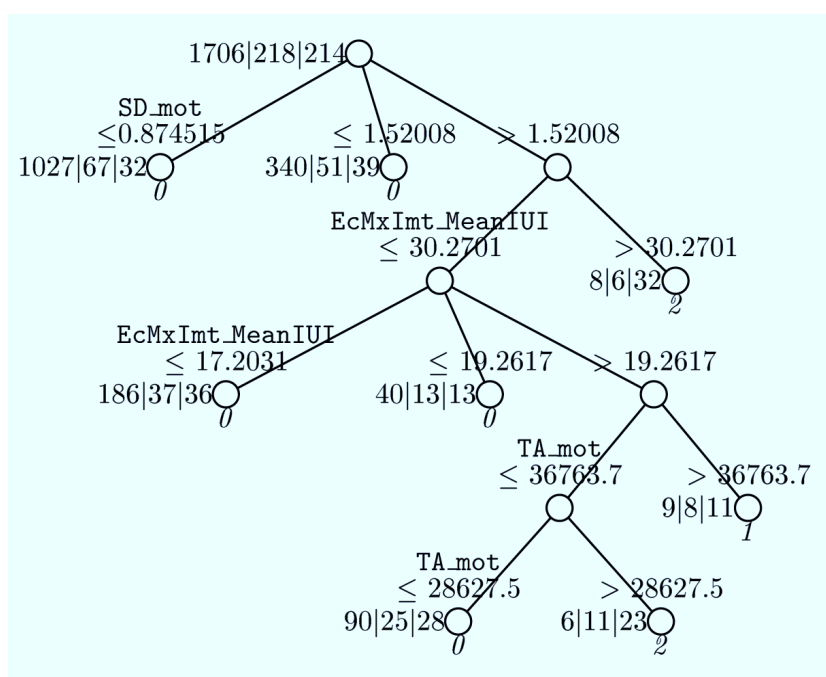
<i>mot prec</i>	mot précédent
<i>mot suiv</i>	mot suivant
<i>UI</i>	unité intonative
<i>UT</i>	unité tonale
<i>% emph</i>	pourcentage d'emphase
<i>coef emph1</i>	coefficient d'emphase n°1
<i>coef emph2</i>	coefficient d'emphase n°2
<i>coef emph3</i>	coefficient d'emphase n°3
<i>coef emph4</i>	coefficient d'emphase n°4
<i>coef emph5</i>	coefficient d'emphase n°5
<i>% degr emph</i>	pourcentage de degré d'emphase
<i>coef degr emph</i>	coefficient de degré d'emphase
<i>nb syll mot</i>	nombre de syllabes du mot
<i>nb syll avt acc</i>	nombre de syllabes avant l'accent primaire du mot
<i>debut mot</i>	temps de début du mot
<i>fin mot</i>	temps de fin du mot
<i>dur_reelle mot</i>	durée réelle du mot
<i>dur_predite mot</i>	durée prédite du mot
<i>allong ms mot</i>	allongement du mot en millisecondes
<i>allong % mot</i>	allongement du mot en pourcentage
<i>MinF0 mot</i>	valeur minimale de F0 sur le mot
<i>TMinF0 mot</i>	temps où apparaît la valeur minimale de F0 sur le mot
<i>MaxF0 mot</i>	valeur maximale de F0 sur le mot
<i>TMaxF0 mot</i>	temps où apparaît la valeur maximale de F0 sur le mot
<i>Mean mot</i>	valeur moyenne de la F0 sur le mot
<i>SD mot</i>	écart-type pour la F0 par rapport à la moyenne du mot
<i>MAS mot</i>	pente moyenne absolue pour la F0 sur le mot
<i>MASWOJ mot</i>	pente moyenne absolue sans saut d'octave pour la F0 sur le mot
<i>RF mot</i>	paramètre indiquant si la F0 sur le mot est en chute (Fall) ou en montée (Rise)
<i>MinF0 UT</i>	valeur minimale de F0 sur l'UT
<i>TMinF0 UT</i>	temps où apparaît la valeur minimale de F0 sur l'UT
<i>MaxF0 UT</i>	valeur maximale de F0 sur l'UT
<i>TMaxF0 UT</i>	temps où apparaît la valeur maximale de F0 sur l'UT
<i>Mean UT</i>	valeur moyenne de la F0 sur l'UT
<i>SD UT</i>	écart-type pour la F0 par rapport à la moyenne de l'UT
<i>MAS UT</i>	pente moyenne absolue pour la F0 sur l'UT
<i>MASWOJ UT</i>	pente moyenne absolue sans saut d'octave pour la F0 sur l'UT
<i>MinF0 UI</i>	valeur minimale de F0 sur l'UI
<i>TMinF0 UI</i>	temps où apparaît la valeur minimale de F0 sur l'UI

<i>MaxF0_UI</i>	valeur maximale de F0 sur l'UI
<i>TMaxF0_UI</i>	temps où apparaît la valeur maximale de F0 sur l'UI
<i>Mean_UI</i>	valeur moyenne de la F0 sur l'UI
<i>SD_UI</i>	écart-type pour la F0 par rapport à la moyenne de l'UI
<i>MAS_UI</i>	pente moyenne absolue pour la F0 sur l'UI
<i>MASWOJ_UI</i>	pente moyenne absolue sans saut d'octave pour la F0 sur l'UI
<i>EcMxF0mt_MeanUI</i>	écart entre la valeur maximale de F0 sur le mot et la valeur moyenne de F0 sur l'UI
<i>EcMxF0UT_MeanUI</i>	écart entre la valeur maximale de F0 sur l'UT et la valeur moyenne de F0 sur l'UI
<i>EcMxF0mot_UI</i>	écart entre la valeur maximale de F0 sur le mot et la valeur maximale de F0 sur l'UI
<i>EcMxF0UT_UI</i>	écart entre la valeur maximale de F0 sur l'UT et la valeur maximale de F0 sur l'UI
<i>EcMnmt_MeanF0UI</i>	écart entre la valeur minimale de F0 sur le mot et la valeur moyenne de F0 sur l'UI
<i>EcMnUT_MeanF0UI</i>	écart entre la valeur minimale de F0 sur l'UT et la valeur moyenne de F0 sur l'UI
<i>EcMinF0mot_UI</i>	écart entre la valeur minimale de F0 sur le mot et la valeur minimale de F0 sur l'UI
<i>EcMinF0UT_UI</i>	écart entre la valeur minimale de F0 sur l'UT et la valeur minimale de F0 sur l'UI
<i>EcMeanF0mot_UI</i>	écart entre la valeur moyenne de F0 sur le mot et la valeur moyenne de F0 sur l'UI
<i>EcMeanF0UT_UI</i>	écart entre la valeur moyenne de F0 sur l'UT et la valeur moyenne de F0 sur l'UI
<i>EcMASmot_UI</i>	écart entre la pente moyenne absolue de F0 sur le mot et la pente moyenne absolue de F0 sur l'UI
<i>EcMASUT_UI</i>	écart entre la pente moyenne absolue de F0 sur l'UT et la pente moyenne absolue de F0 sur l'UI
<i>EcMASWOJmot_UI</i>	écart entre la pente moyenne absolue sans saut d'octave de F0 sur le mot et la pente moyenne absolue sans saut d'octave de F0 sur l'UI
<i>EcMASWOJUT_UI</i>	écart entre la pente moyenne absolue sans saut d'octave de F0 sur l'UT et la pente moyenne absolue sans saut d'octave de F0 sur l'UI
<i>MinInt_mot</i>	valeur minimale de l'intensité sur le mot
<i>TMinInt_mot</i>	temps où apparaît la valeur minimale de l'intensité sur le mot
<i>MaxInt_mot</i>	valeur maximale de l'intensité sur le mot
<i>TMaxInt_mot</i>	temps où apparaît la valeur maximale de l'intensité sur le mot
<i>MeanInt_mot</i>	valeur moyenne de l'intensité sur le mot
<i>SDInt_mot</i>	écart-type pour l'intensité par rapport à la moyenne sur le mot
<i>MinInt_UT</i>	valeur minimale de l'intensité sur l'UT
<i>TMinInt_UT</i>	temps où apparaît la valeur minimale de l'intensité sur l'UT
<i>MaxInt_UT</i>	valeur maximale de l'intensité sur l'UT
<i>TMaxInt_UT</i>	temps où apparaît la valeur maximale de l'intensité sur l'UT
<i>MeanInt_UT</i>	valeur moyenne de l'intensité sur l'UT
<i>SDInt_UT</i>	écart-type pour l'intensité par rapport à la moyenne sur l'UT
<i>MinInt_UI</i>	valeur minimale de l'intensité sur l'UI

<i>TMinInt_UI</i>	temps où apparaît la valeur minimale de l'intensité sur l'UI
<i>MaxInt_UI</i>	valeur maximale de l'intensité sur l'UI
<i>TMaxInt_UI</i>	temps où apparaît la valeur maximale de l'intensité sur l'UI
<i>MeanInt_UI</i>	valeur moyenne de l'intensité sur l'UI
<i>SDInt_UI</i>	écart-type pour l'intensité par rapport à la moyenne sur l'UI
<i>EcMxInt_MeanIUI</i>	écart entre la valeur maximale de l'intensité sur le mot et la valeur moyenne de l'intensité sur l'UI
<i>EcMxIUT_MeanIUI</i>	écart entre la valeur maximale de l'intensité sur l'UT et la valeur moyenne de l'intensité sur l'UI
<i>EcMinInt_MeanIUI</i>	écart entre la valeur minimale de l'intensité sur le mot et la valeur moyenne de l'intensité sur l'UI
<i>EcMinIUT_MeanIUI</i>	écart entre la valeur minimale de l'intensité sur l'UT et la valeur moyenne de l'intensité sur l'UI
<i>EcMxIntmot_UI</i>	écart entre la valeur maximale de l'intensité sur le mot et la valeur maximale de l'intensité sur l'UI
<i>EcMxIntUT_UI</i>	écart entre la valeur maximale de l'intensité sur l'UT et la valeur maximale de l'intensité sur l'UI
<i>EcMinIntmot_UI</i>	écart entre la valeur minimale de l'intensité sur le mot et la valeur minimale de l'intensité sur l'UI
<i>EcMinIUT_UI</i>	écart entre la valeur minimale de l'intensité sur l'UT et la valeur minimale de l'intensité sur l'UI
<i>EcMeanIntmot_UI</i>	écart entre la valeur moyenne de l'intensité sur le mot et la valeur moyenne de l'intensité sur l'UI
<i>EcMeanIUT_UI</i>	écart entre la valeur moyenne de l'intensité sur l'UT et la valeur moyenne de l'intensité sur l'UI
<i>TA_mot</i>	amplitude totale sur le mot

Annexe 4b - Arbres de classification : corpus primaire

Groupes 0 à 8 (1SE) :

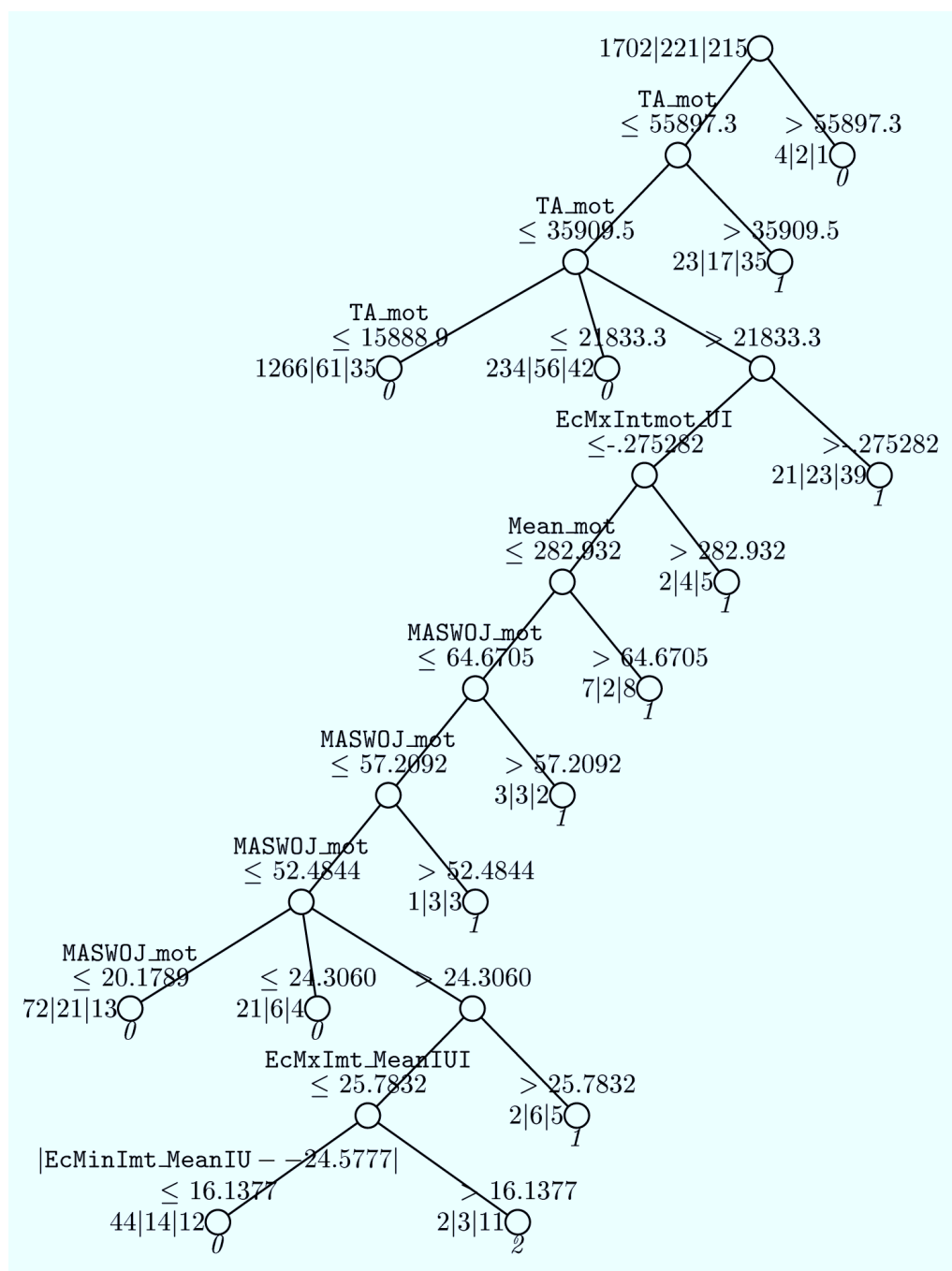


Classification Matrix for Test Data emph-9

	Predicted Class		
	0	1	2
Actual class 0	191	0	5
Actual class 1	19	1	2
Actual class 2	13	1	5

Total obs in Test sample = 237, # correct = 197

Groupe 1 à 9 (ISE):

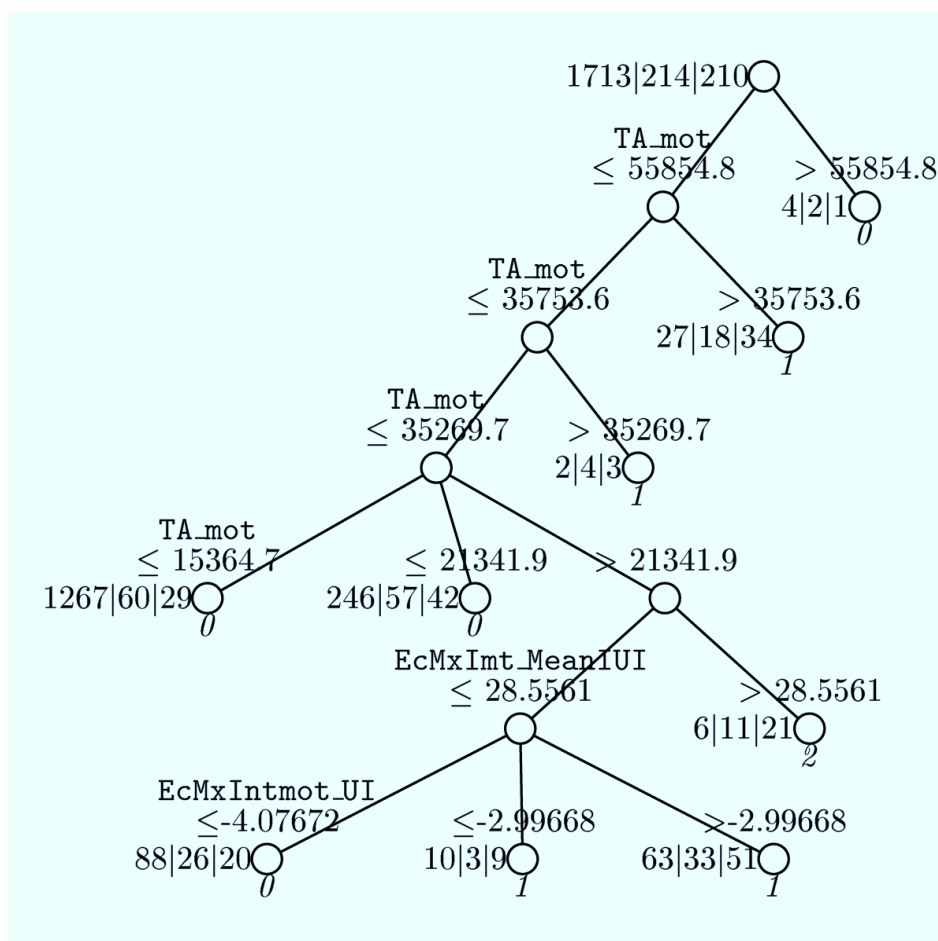


Classification Matrix for Test Data emph-0

Actual class	Predicted Class		
	0	1	2
0	192	8	0
1	13	5	1
2	12	6	0

Total obs in Test sample = 237, # correct = 197

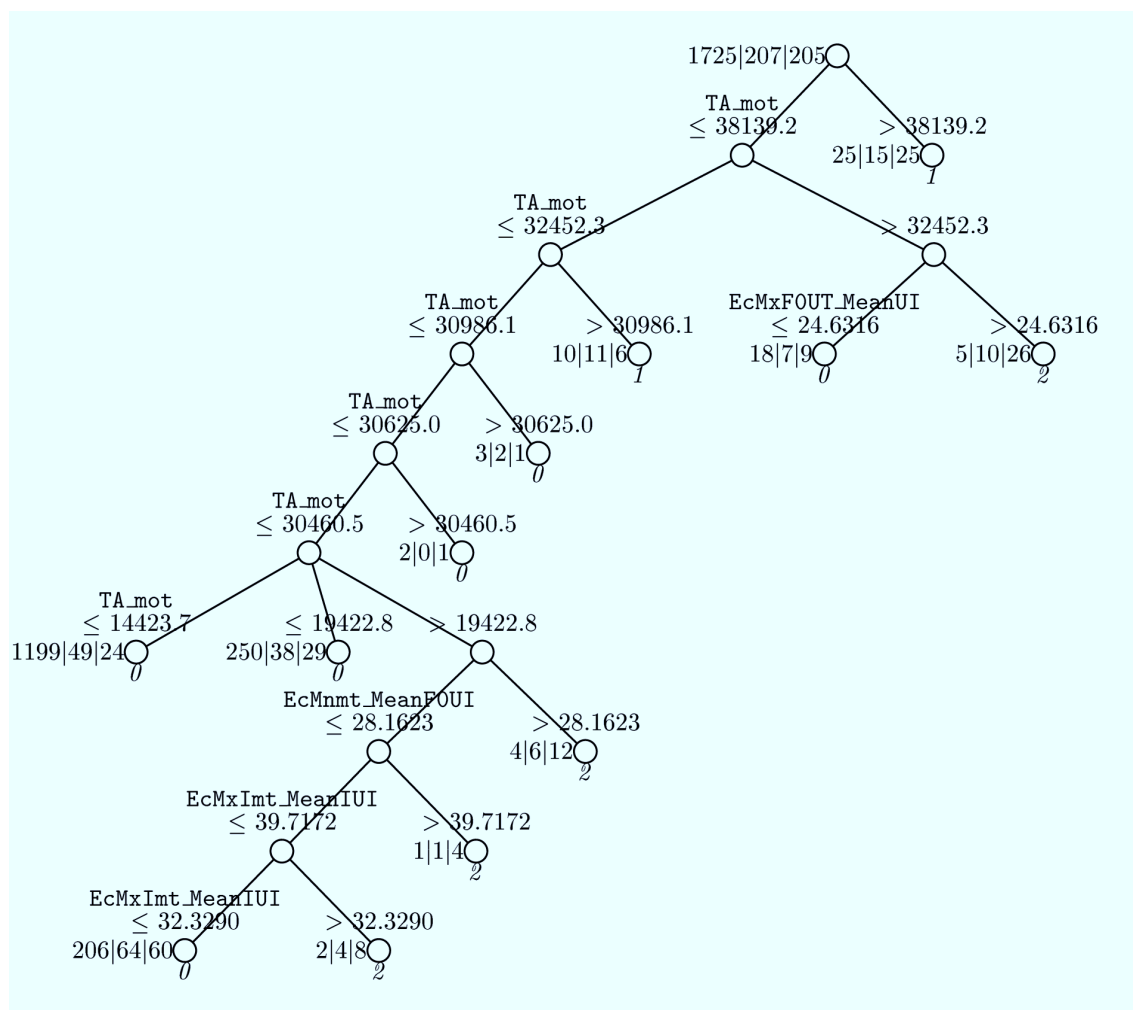
Groupe 2 à 0 (ISE):



Classification Matrix for Test Data emph-1

	Predicted Class		
	0	1	2
Actual class	-----		
0	169	17	3
1	16	7	3
2	10	10	3

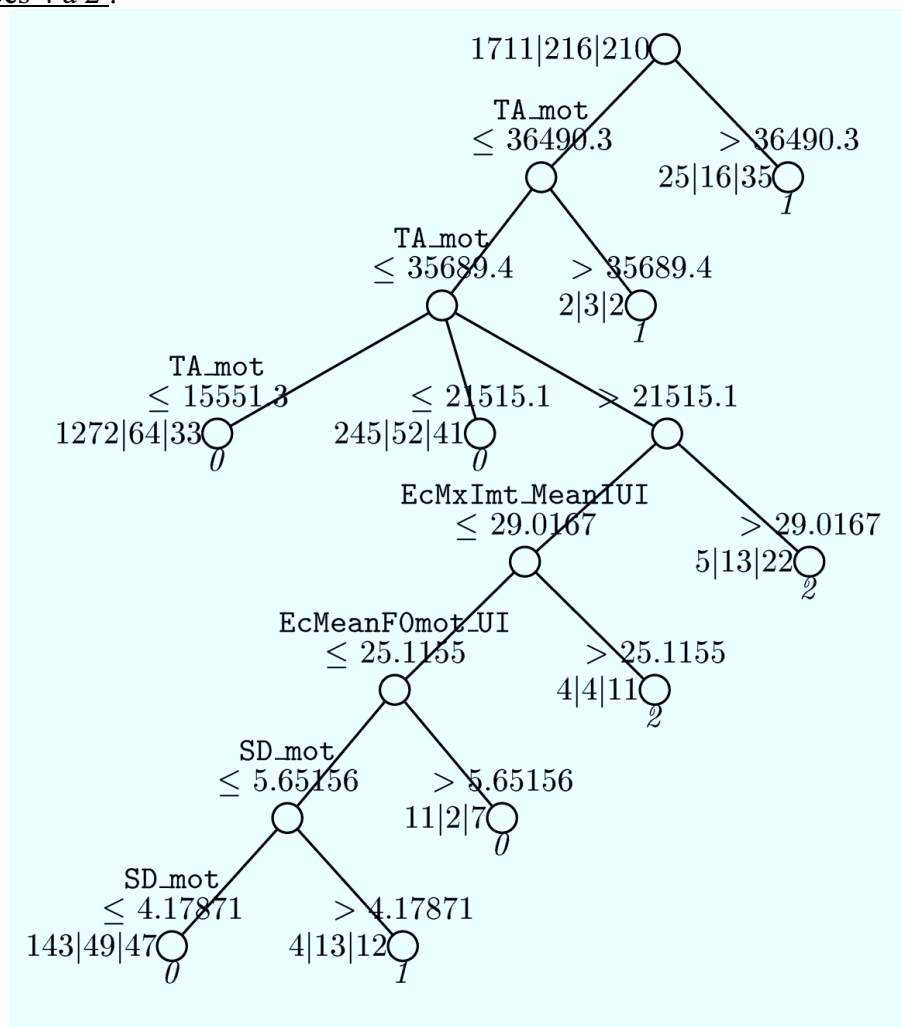
Total obs in Test sample = 238, # correct = 179

Groupe 3 à 1 (OSE):**Classification Matrix for Test Data emph-2**

	Predicted Class		
	0	1	2
Actual class	-----		
0	174	1	2
1	30	1	2
2	19	5	4

Total obs in Test sample = 238, # correct = 179

Groupes 4 à 2 :

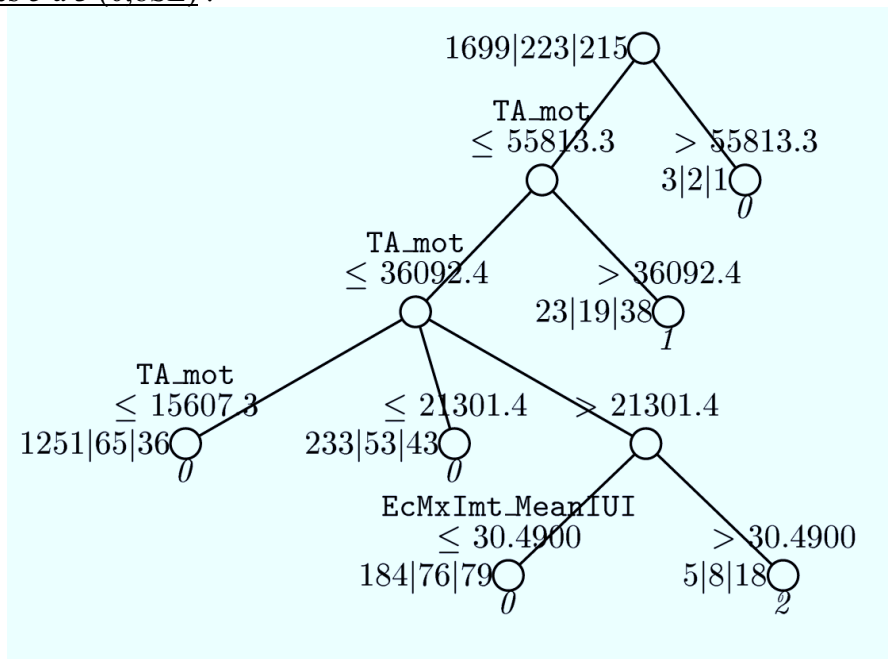


Classification Matrix for Test Data emph-3

	Predicted Class		
	0	1	2
Actual class	-----		
0	180	7	4
1	16	4	4
2	16	5	2

Total obs in Test sample = 238, # correct = 186

Groupe 5 à 3 (0,8SE) :

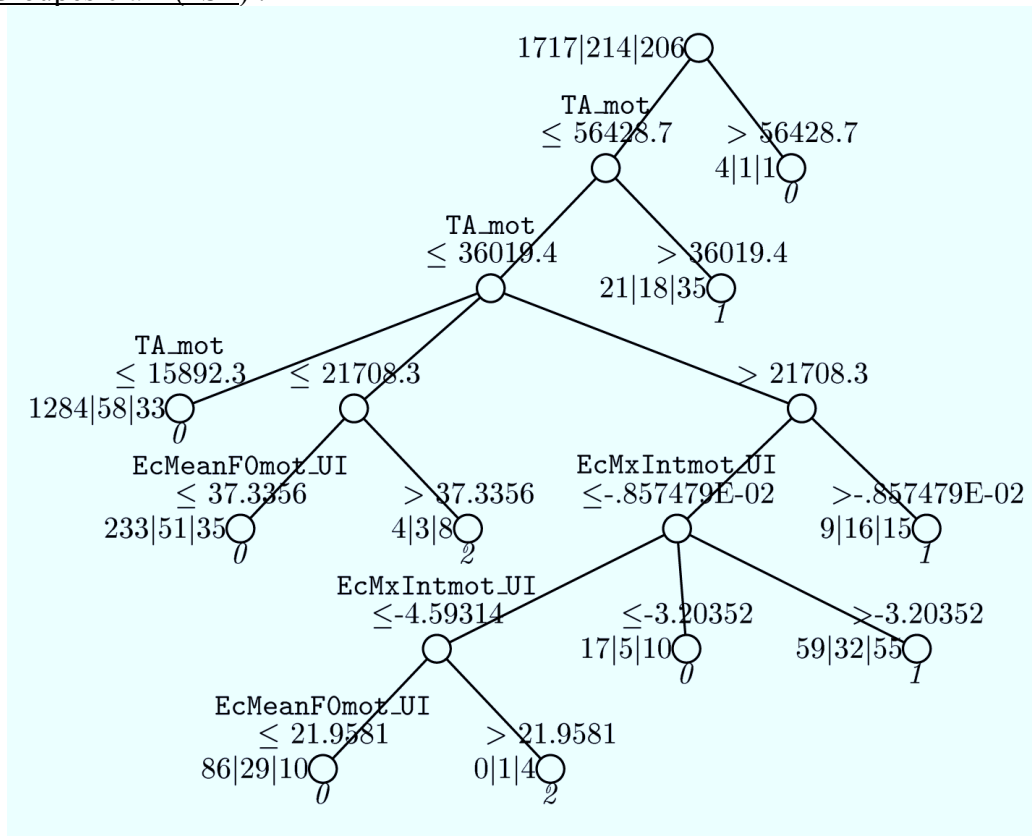


Classification Matrix for Test Data emph-4

Actual class	Predicted Class		
	0	1	2
0	192	11	0
1	8	7	2
2	6	10	2

Total obs in Test sample = 238, # correct = 201

Groupes 6 à 4 (1SE) :

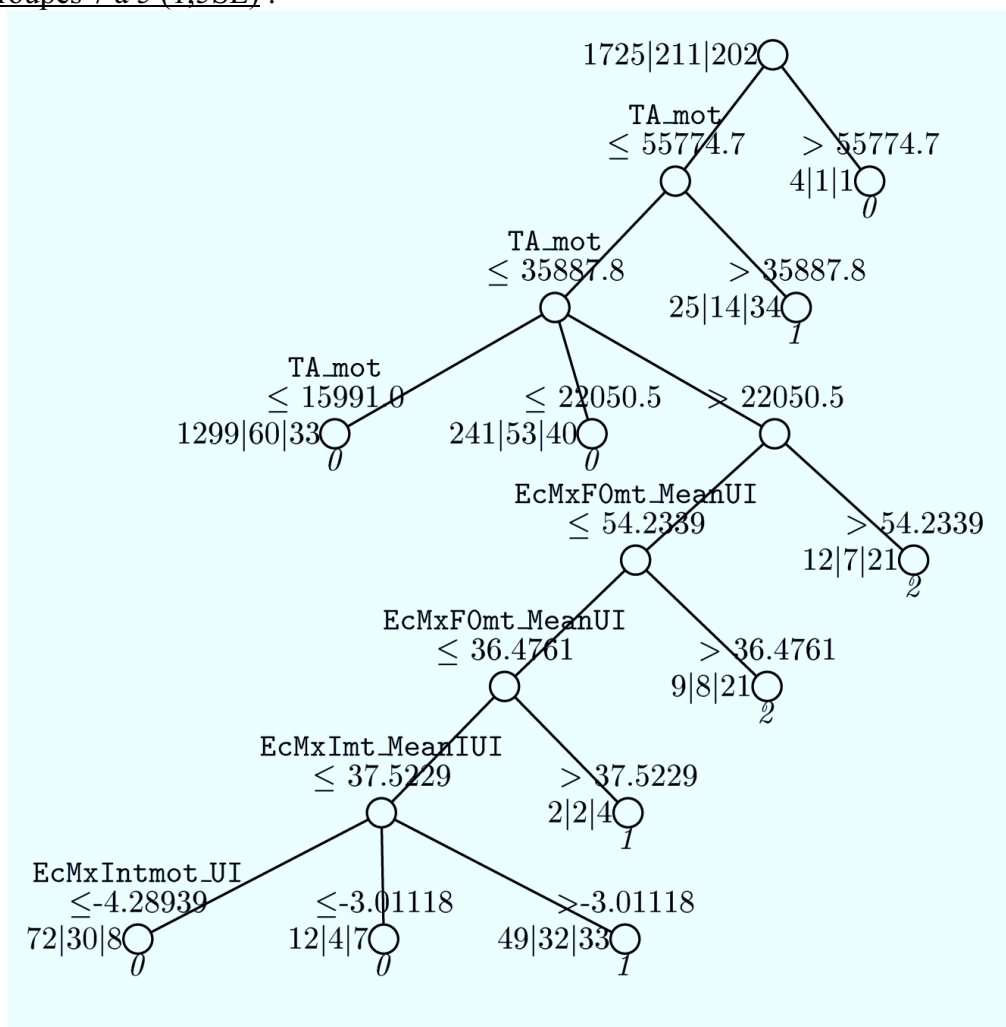


Classification Matrix for Test Data emph-5

	Predicted Class		
	0	1	2
Actual class	-----		
0	172	12	1
1	18	6	2
2	13	13	1

Total obs in Test sample = 238, # correct = 179

Groupes 7 à 5 (1,5SE) :

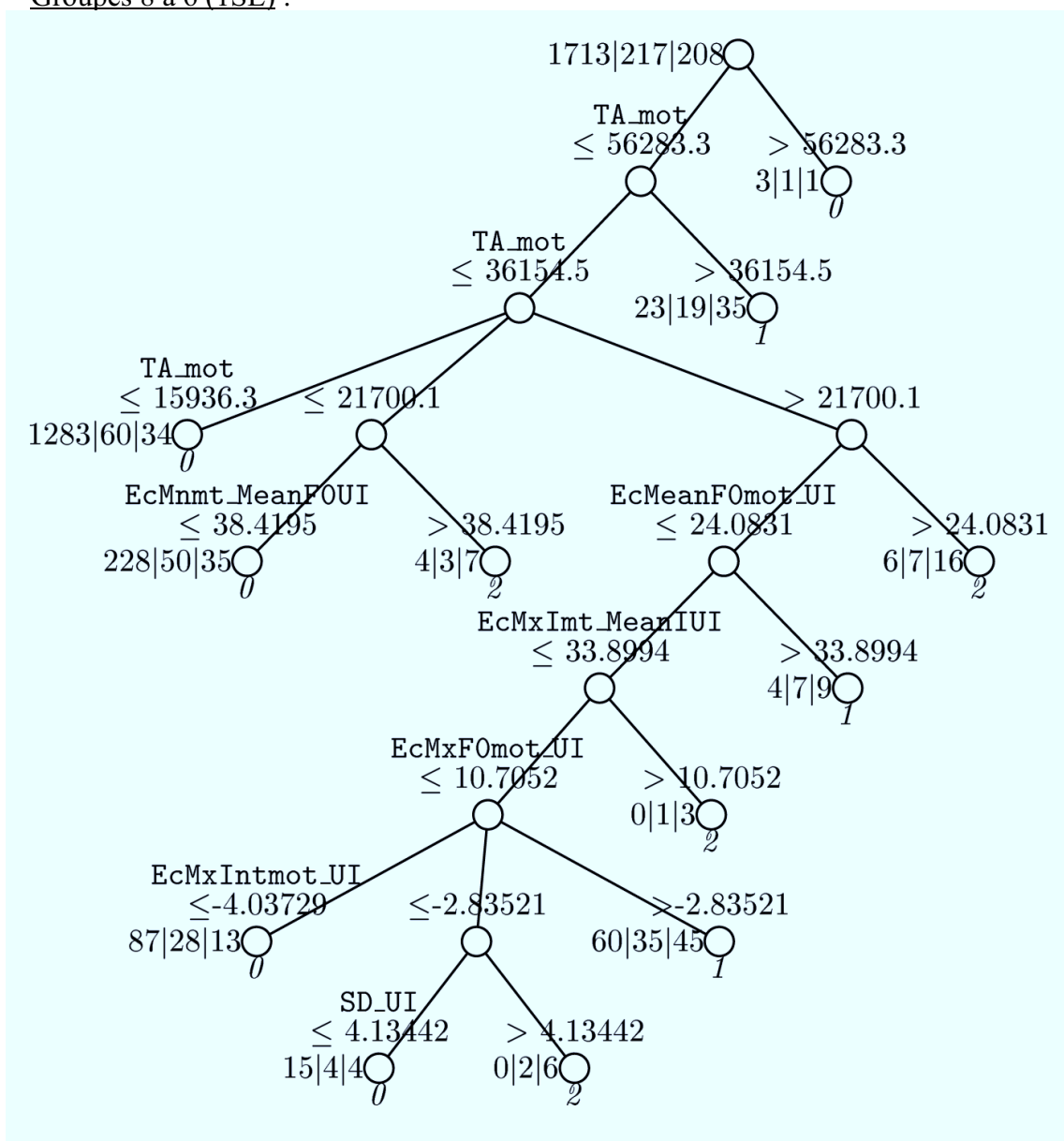


Classification Matrix for Test Data emph-6

	Predicted Class		
	0	1	2
Actual class	-----		
0	166	9	2
1	18	7	4
2	17	9	5

Total obs in Test sample = 237, # correct = 178

Groupes 8 à 6 (1SE) :

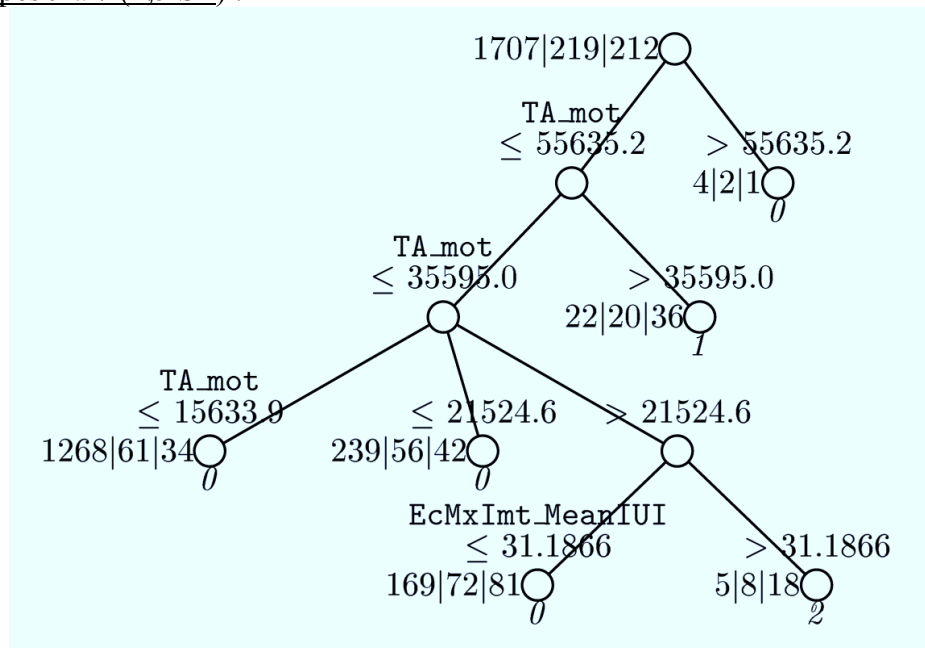


Classification Matrix for Test Data emph-7

	Predicted Class		
	0	1	2
Actual class	-----		
0	182	5	2
1	18	3	2
2	10	12	3

Total obs in Test sample = 237, # correct = 188

Groupes 9 à 7 (1,5 SE) :



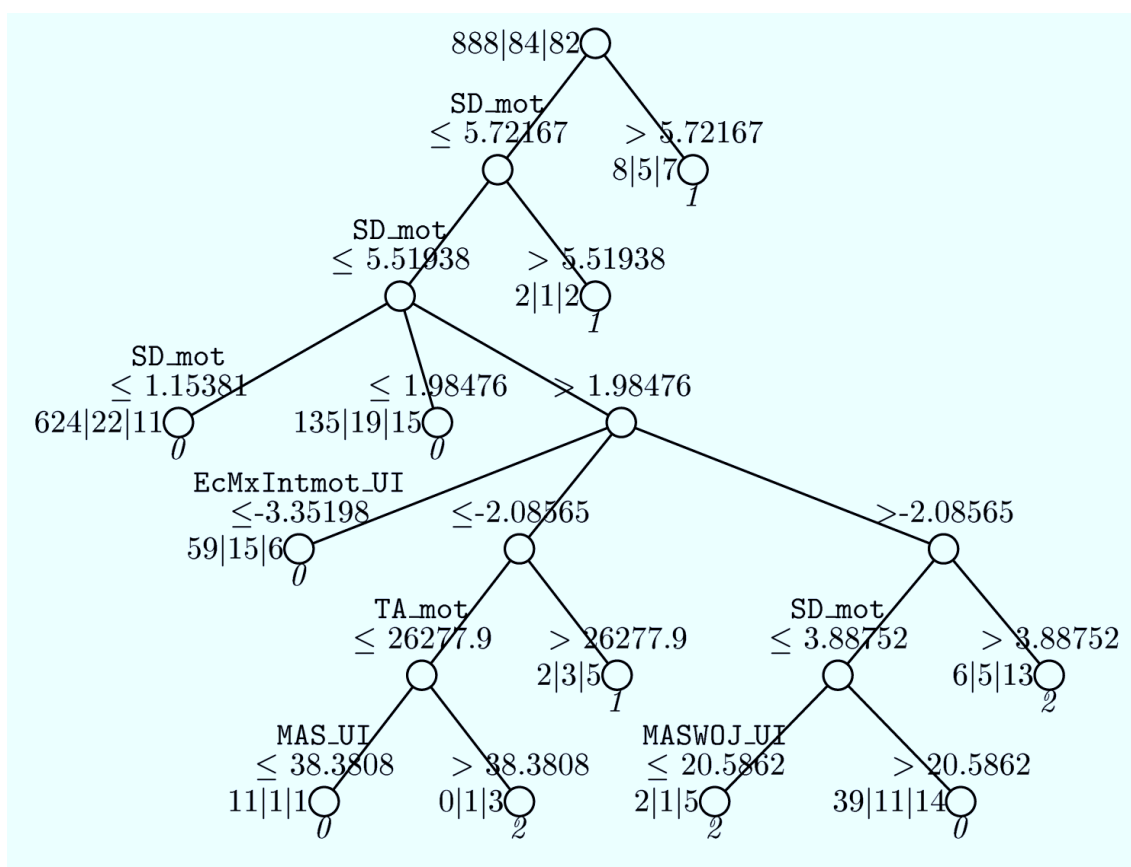
Classification Matrix for Test Data emph-8

	Predicted Class		
	0	1	2
Actual class	-----		
0	189	6	0
1	19	2	0
2	16	4	1
Total obs in Test sample = 237, # correct = 192			

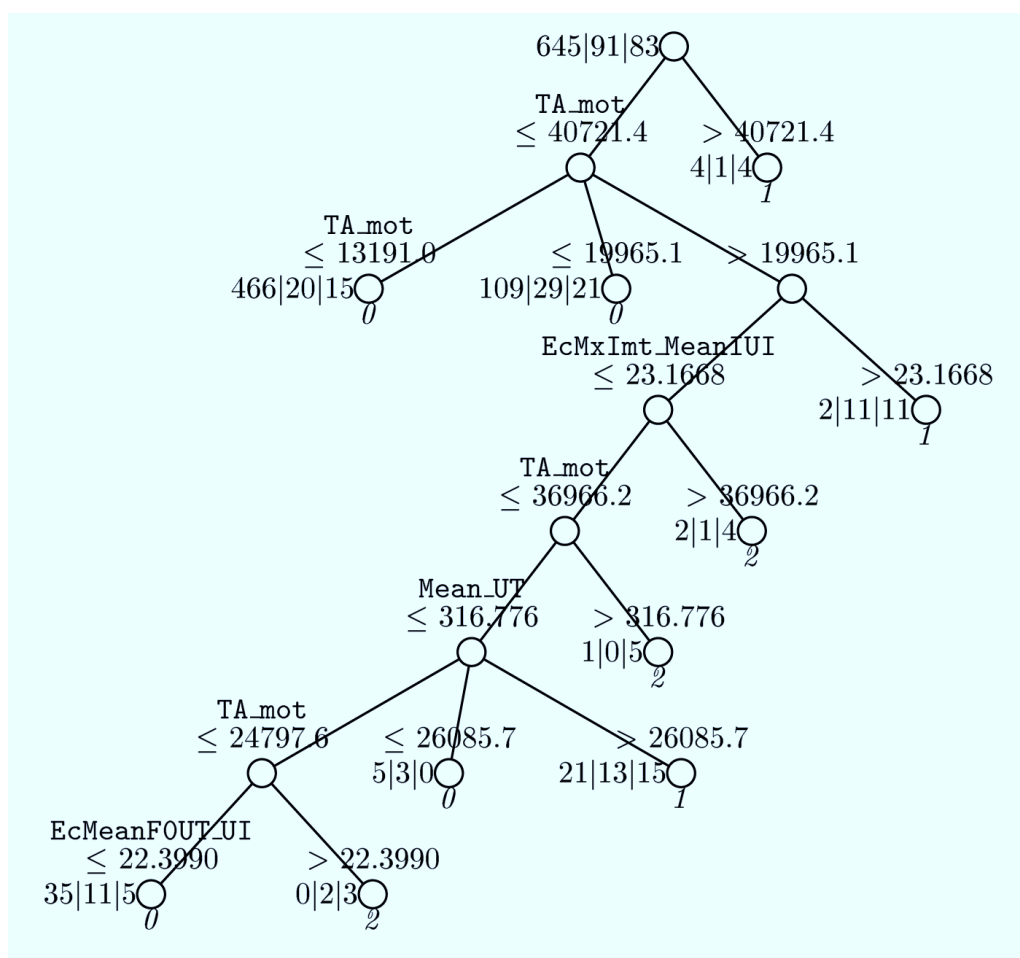


Annexe 4c - Arbres de classification : types de discours

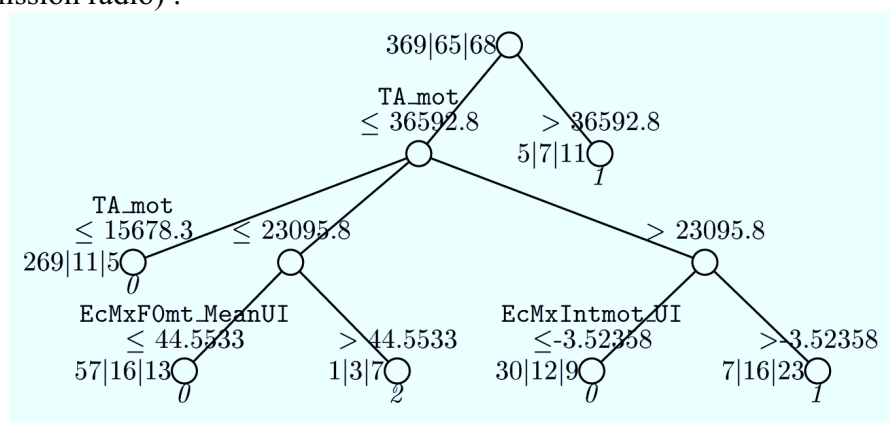
P1 (débat politique) :



P2 (conversation) :

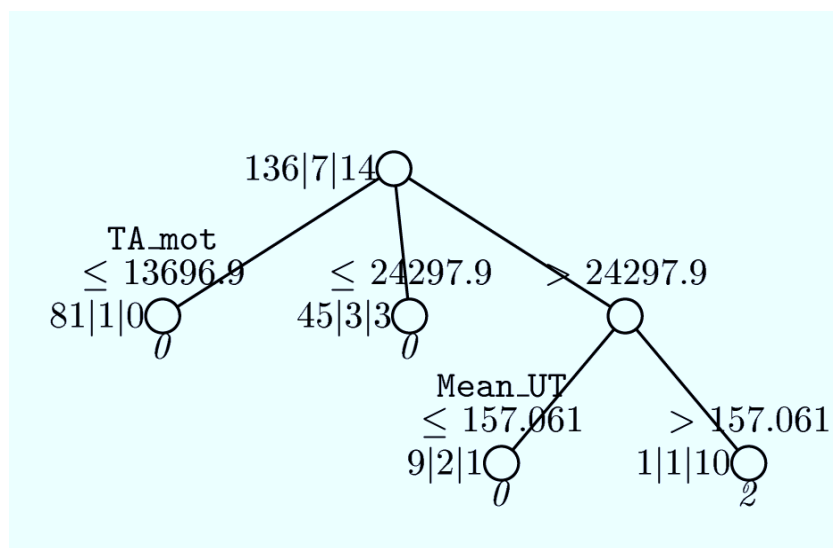


P3 (émission radio) :

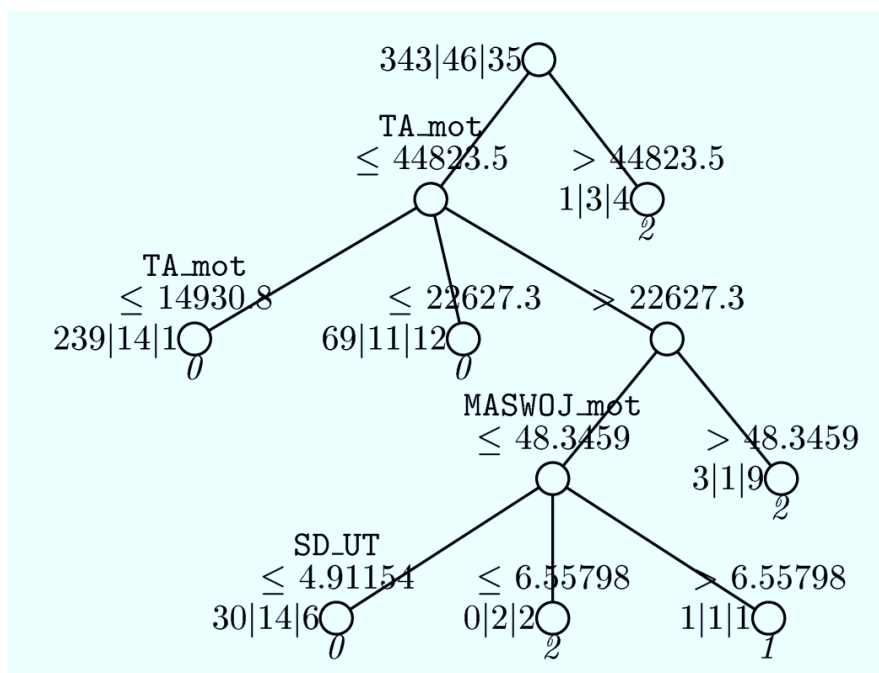


Annexe 4d - Arbres de classification : locuteurs

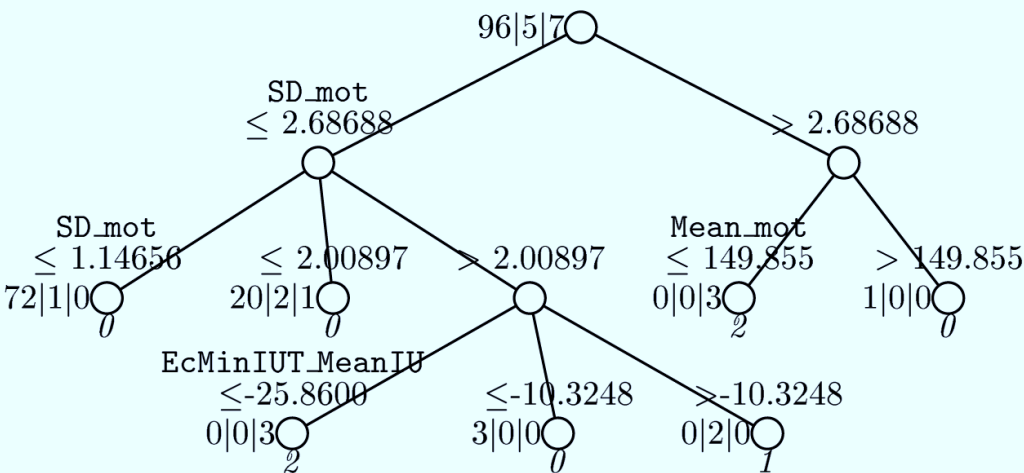
DD :



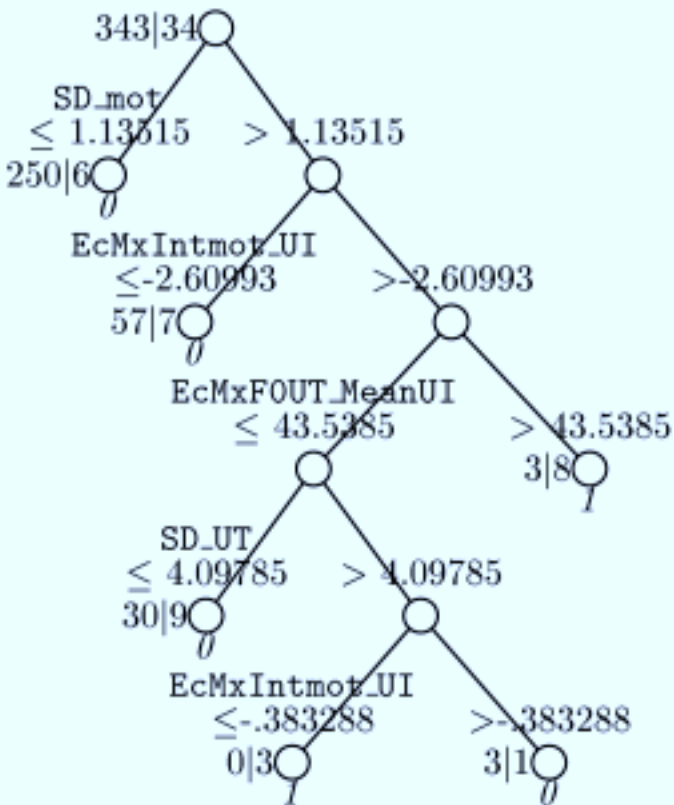
PM :

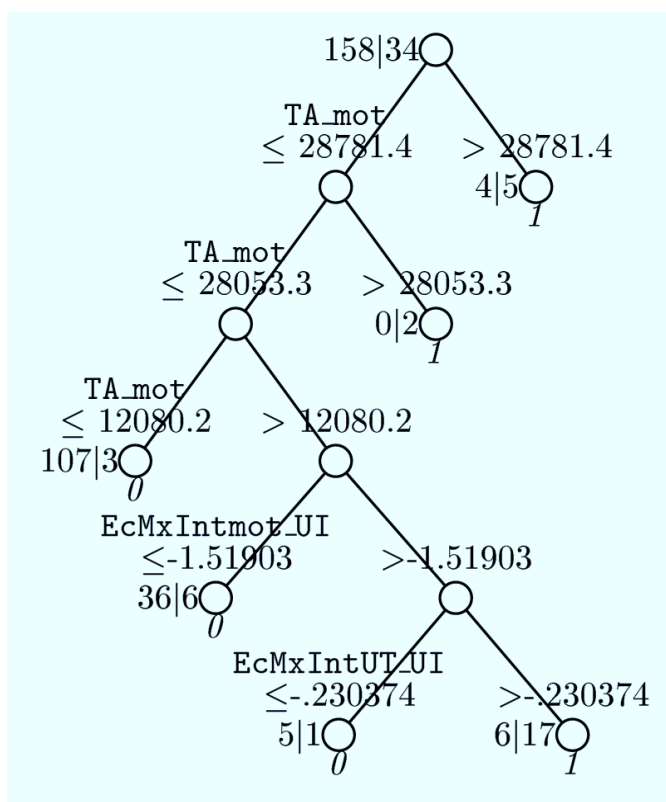
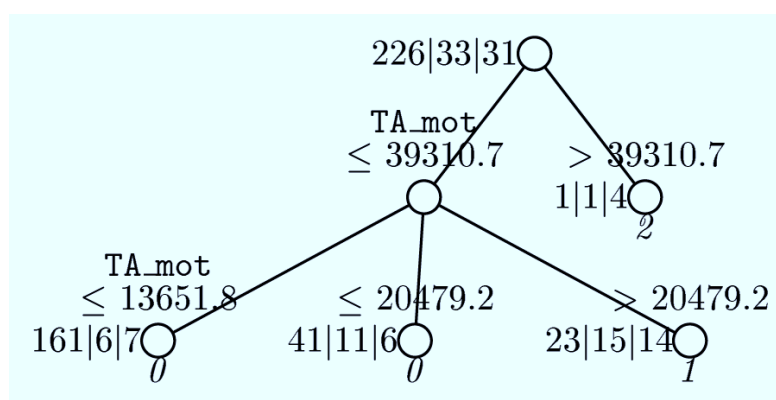


JT:

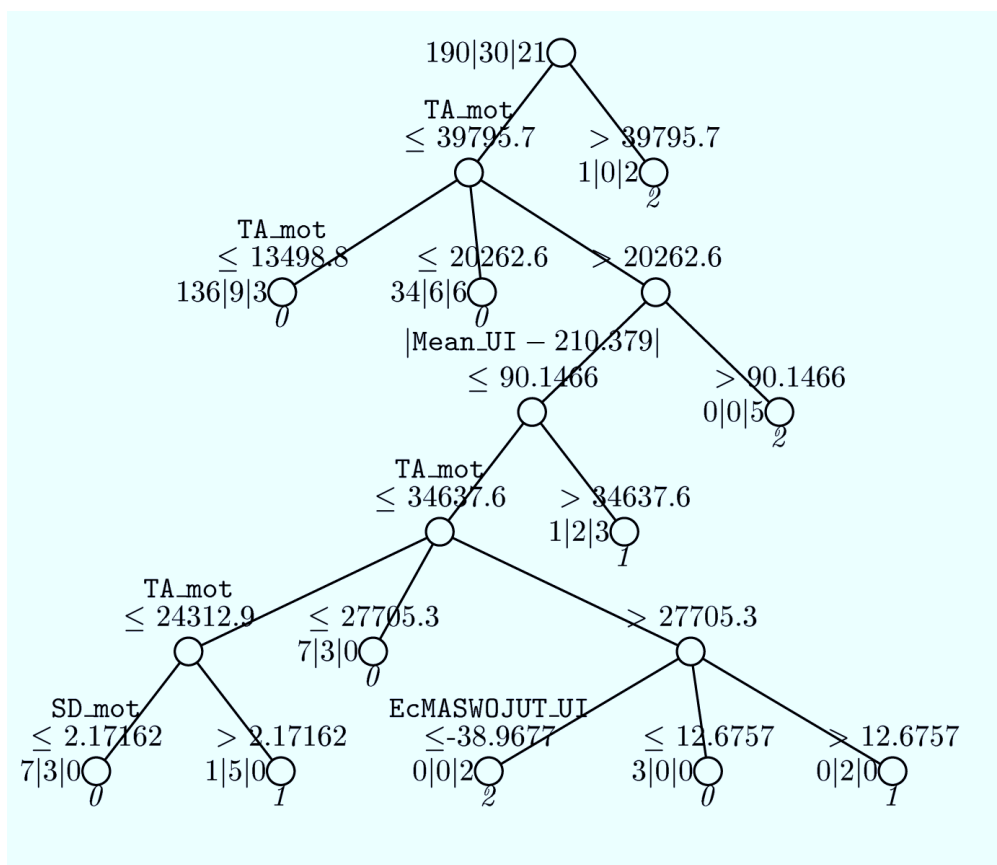


MM :

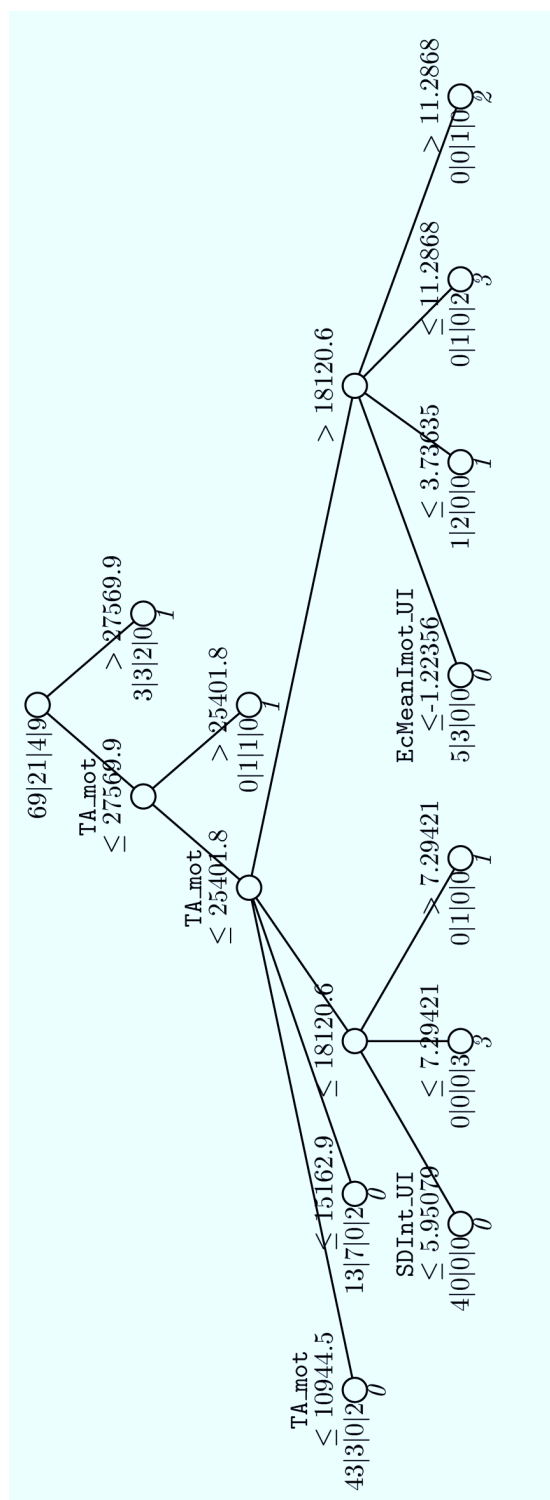


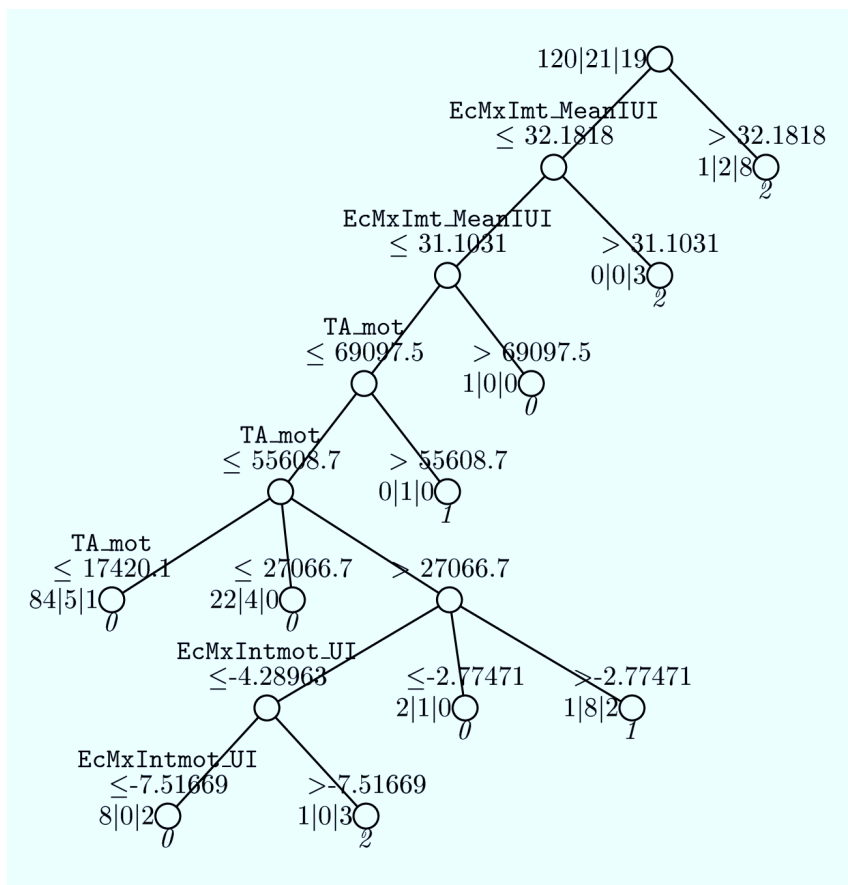
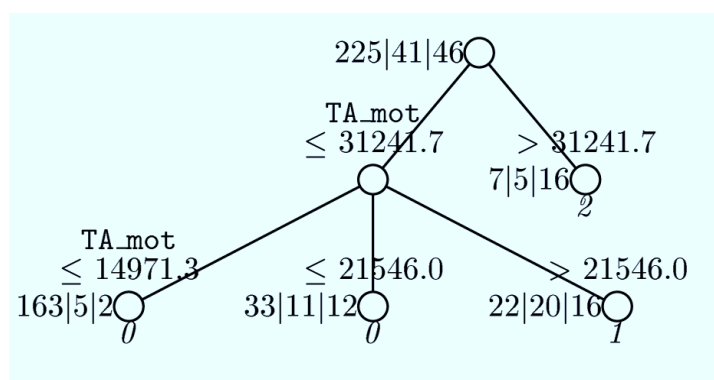
DL :VB :

CB :



JC :



PS :YB :

Annexe 5a - Script ‘mapPhones’

```
#!/usr/bin/perl
# mapPhones      macPerl droplet
#
$version = "1.0";
#
#               author: Daniel Hirst
#               email: daniel.hirst@lpl.univ-aix.fr
#               last modified: 17jun2000
#
# purpose: maps phonemes from one set to another
#
defined in ASCII file "phoneMap.txt"
#
#               input format:
#               phoneme [from SAMPA set A] tab
#               rest_of_line
#
#               output format:
#               phoneme [from SAMPA set B] tab
#               rest_of_line
#
# name of droplet ***mac specific***
#
($droplet_name = $0) =~ s/^\.*://;

$phoneMap = "phoneMap.txt";
open(MAP, $phoneMap) or die ("Could not find $phoneMap");

while (<MAP>) {
    chomp;
# ignore comments (;) or empty lines
    if (($_ !~ /;/) && ($_ =~ /. /)) {
        ($phonemeA, $phonemeB) = split (/t/, $_);
        $map{$phonemeA} = $phonemeB;
    }
}
close (MAP);

$infile = $ARGV[0];
open (IN, $infile) or die("cannot open $infile");
open(IN, $infile) or die ("Could not find $infile");

## remove any extension with "." and add extension in $outtext

($outfile = $infile) =~ s/\.\/_B\./;
```

```

open (OUT, ">$outfile");

# names of input and output files
($infile_name = $infile) =~ s/^.*://;
($outfile_name = $outfile) =~ s/^.*://;

$now = localtime;

print ("; $outfile_name\n");
print ("; $outfile\n");
print ("; converted by $droplet_name on $now\n");
print ("; from $infile_name\n\n");

while (<IN>) {
    chomp;
    #keep comments (;) or empty lines
    if (($_ =~ /;/) || ($_ !~ /./)) {
        print OUT ("$_\n");
    } else {
        ($phoneme, $rest) = split (/[\t]/, $_,
2);
        if (defined ($map{$phoneme})) {
            print OUT
("$map{$phoneme}\t$rest\n");
        } else {die ("Undefined mapping for
phoneme [$phoneme]\n");}
        # input text can be tab or whitespace separated
    }
}
close (IN);
close (OUT);

# version history
#
# 17jun2000      1.0

```

Annexe 5b - Test perceptif sur signaux manipulés : auditeurs

Groupe 1 :

MM : étudiante, Manchester, Engl, 23
HG : étudiante, London, Engl, 23
DB : étudiant, Engl, 23
CV : mère au foyer, Birmingham, Engl, 38
PN : étudiante, S.E., England, 21
HM : étudiant, Bristol, Engl, 20
AW : étudiante, Ali Westmoreland, South Engl, 22
NO : étudiant, Athlone, Ireland, 25
AB : homme, Lancashire, Engl, 37
RM : avocate, S.E. England, 53

Groupe 2 :

CL : étudiant, Leitrim, Ireland, North West, 24
VH : enseignante, S.W. England, 44
TB : étudiante, S.E. Engl, 24
ML : lectrice, Midlands Ireland, 23
JO : étudiante, London, 21
AC : étudiante, Manchester, 24
UB : graphiste, Ursula Brain, S.E. Engl, 27
MB : retraité, Cornwall, 74
CL : mère au foyer, Cornwall, 47
CM : militaire, N.W. England, 36

Annexe 5c - Test perceptif sur signaux manipulés : instructions

Perception experiment based on synthesized speech.

I'm writing a PhD on the acoustic and prosodic correlates of emphasis. I need native English speakers to tell me what they feel is pronounced emphatically. What I mean by emphatic is NOT NEUTRAL, uttered with SPECIAL INVOLVEMENT, made PROMINENT in an emphatic (exaggerated) way.

So you are going to hear sentences which you can read on the sheet of paper, and I would like you to underline the passages, words or syllables which which you feel are being said in an emphatic way.

This perception experiment is based on synthesized manipulated speech.

You are first going to hear sentences uttered in a language which is unknown to you (please, write it down on the paper if you recognize that language). You will hear each sentence twice. Please, underline the emphatic passages, or words, if any.

Then you will hear sentences in English, each sentence repeated twice as well. Please, do the same : underline the emphatic passages, words or syllables, if any.

Not all the passages you are going to hear are necessarily emphatic.

Any comments are welcome and will probably be helpful. THANK YOU !

Surname (you can just write the initial if you prefer) :

Name (you can just write the initial if you prefer) :

Age :

Country and region you come from :

PART ONE (unknown language)

- * Zas az a tasasr tada aza jaza za tansra ad sratnd ta razan
- * And a, latan at at, a sant ja satnra daan ta saz a data tans as sna an trasnas da
- * Za z a tasatra sansadr tarasa jat za tansazataz tadanat za tansazataz dazanant an za tansazataz tata an za sas a tananz
- * An zat az zat ja daan ta jat ta sa sa zas adjanant dazarats an a sandr taransa
- * Jat ja jan ta da az atsra dat an za, an ratrazant a tantra z antrast
- * Na ats sra tradras, dat za a natan tradras
- * Da dad ja da zat rast jar az jar
- * An at jaz adsaratra srazan
- * An za tarat sad jarz za tan dadra dasrans tjan nasaj an an an randn
- * A sant tatr sa ran a landjad, das da tatan at at, ja na az adarts zan za tradra da tat at ar sats a nastats
- * At s nat azan dan dastast tar zas ratat tan at tratara
- * A jad ja sadanastat zaarans az a trazat asja

PART TWO (English) GROUP 1

You will hear each sentence twice. There are pairs of sentences given at random.

- * or would you say domestic violence is a private issue
- * there's a perfectly sensible policy which the Conservative cabinet the Conservative government and the Conservative party and the House of Commons

* what we want to do is actually get in there and represent our country's interests

* looking at it I think we're certainly going to have a better chance of snow on Christmas Day

* It's not even been discussed till this report came out properly

* now it's slow progress but they are making progress

* or would you say domestic violence is a private issue

* and the pilot said well there's a ten-degree difference between Marseille and and and London

* and it was absolutely freezing

* now it's slow progress but they are making progress

* Did you do that last year as well ?

* This is a kerfuffle today over whether the Chancellor had threatened to resign

* there's a perfectly sensible policy which the Conservative cabinet the Conservative government and the Conservative party and the House of Commons

* Did you do that last year as well ?

* and that is that we're going to wait to see how this argument develops on a single currency

* looking at it I think we're certainly going to have a better chance of snow on Christmas Day

* It's not even been discussed till this report came out properly

* I think people who learn a language just by picking it up you know as adults then they probably do pick up all sorts of mistakes.

- * what we want to do is actually get in there and represent our country's interests
- * and that is that we're going to wait to see how this argument develops on a single currency
- * and it was absolutely freezing
- * This is a kerfuffle today over whether the Chancellor had threatened to resign
- * and the pilot said well there's a ten-degree difference between Marseille and and and London
- * I think people who learn a language just by picking it up you know as adults then they probably do pick up all sorts of mistakes.

PART TWO (English) GROUP 2

You will hear each sentence twice. There are pairs of sentences given atrandom.

- * now it's slow progress but they are making progress
- * what we want to do is actually get in there and represent our country's interests
- * This is a kerfuffle today over whether the Chancellor had threatened to resign
- * It's not even been discussed till this report came out properly
- * looking at it I think we're certainly going to have a better chance of snow on Christmas Day
- * Did you do that last year as well ?
- * It's not even been discussed till this report came out properly
- * There's a perfectly sensible policy which the Conservative cabinet the Conservative government and the Conservative party and the House of Commons
- * And that is that we're going to wait to see how this argument develops

on a single currency

* And it was absolutely freezing

* Did you do that last year as well ?

* and the pilot said well there's a ten-degree difference between Marseille and and and London

* I think people who learn a language just by picking it up you know as adults then they probably do pick up all sorts of mistakes.

* or would you say domestic violence is a private issue

* now it's slow progress but they are making progress

* This is a kerfuffle today over whether the Chancellor had threatened to resign

* I think people who learn a language just by picking it up you know as adults then they probably do pick up all sorts of mistakes.

* what we want to do is actually get in there and represent our country's interests

* and that is that we're going to wait to see how this argument develops on a single currency

* and it was absolutely freezing

* or would you say domestic violence is a private issue

* and the pilot said well there's a ten-degree difference between Marseille and and and London

* looking at it I think we're certainly going to have a better chance of snow on Christmas Day

* there's a perfectly sensible policy which the Conservative cabinet the Conservative government and the Conservative party and the House of Commons

Annexe 5d - Test perceptif sur signaux manipulés : résultats

Ce tableau montre le pourcentage d'auditeurs qui ont pointé le mot emphatique dans les différents stimulus.

Segment	mots	normal	normal synth	manip durée	manip F0	manip pauses	jastradanz
P1.1S01	This	0	0	10	0	10	0
P1.1S01	is	0	0	0	0	0	10
P1.1S01	a	0	0	0	0	20	25
P1.1S01	kerfuffle	83,3	50	30	40	80	65
P1.1S01	today	27,7	0	0	0	10	5
P1.1S01	over	0	0	0	0	0	0
P1.1S01	whether	0	0	0	0	0	0
P1.1S01	the	0	0	0	0	0	0
P1.1S01	Chancellor	5,5	30	50	20	30	10
P1.1S01	had	0	0	0	0	0	5
P1.1S01	threatened	11,1	10	20	20	30	15
P1.1S01	to	0	0	0	0	0	0
P1.1S01	resign	83,3	50	40	30	30	65
P1.3S01	looking	11,1	0	0	10	20	15
P1.3S01	at	0	0	0	10	0	55
P1.3S01	it	0	0	0	0	0	0
P1.3S01	I	0	0	0	0	0	0
P1.3S01	think	0	10	10	0	0	0
P1.3S01	we're	0	0	0	0	10	0
P1.3S01	certainly	5,5	0	0	10	40	5
P1.3S01	going	0	0	0	0	0	0
P1.3S01	to	0	0	0	0	0	10
P1.3S01	have	0	0	0	0	0	0
P1.3S01	a	0	0	0	0	0	0
P1.3S01	better	11,1	0	0	0	20	0
P1.3S01	chance	0	0	0	10	0	5
P1.3S01	of	0	0	0	0	0	5
P1.3S01	snow	94,4	100	60	90	80	15
P1.3S01	on	0	0	0	0	0	10
P1.3S01	Christmas	77,7	30	40	20	60	90
P1.3S01	Day	38,8	40	30	50	40	10
P1.4S04	there's	5,5	0	0	0	0	10

P1.4S04	a	5,5	0	0	0	0	15
P1.4S04	perfectly	72,2	80	30	20	80	15
P1.4S04	sensible	22,2	10	20	10	10	5
P1.4S04	policy	16,6	20	10	20	0	20
P1.4S04	which	0	0	0	0	0	0
P1.4S04	the	0	0	0	0	0	0
P1.4S04	Conservative	11,1	30	30	20	40	10
P1.4S04	cabinet	94,4	50	80	30	40	10
P1.4S04	the	0	0	0	0	0	0
P1.4S04	Conservative	16	20	20	20	40	15
P1.4S04	government	88,8	40	20	30	50	15
P1.4S04	and	5,5	0	0	0	0	0
P1.4S04	the	5,5	0	0	0	0	0
P1.4S04	Conservative	16,6	20	20	20	30	5
P1.4S04	party	44,4	40	20	30	30	0
P1.4S04	and	0	0	0	0	0	10
P1.4S04	the	0	0	0	0	0	5
P1.4S04	House	0	0	0	10	20	0
P1.4S04	of	0	0	0	0	10	0
P1.4S04	Commons	0	0	0	10	20	10
P1.4S05	and	0	0	0	0	10	5
P1.4S05	that	0	20	10	40	60	5
P1.4S05	is	0	0	30	0	20	10
P1.4S05	that	0	10	10	0	10	15
P1.4S05	we're	0	0	10	0	10	5
P1.4S05	going	22,2	0	0	0	10	0
P1.4S05	to	11,1	0	0	0	20	0
P1.4S05	wait	77,7	10	10	0	80	10
P1.4S05	to	33,3	0	0	0	20	5
P1.4S05	see	77,7	30	10	0	80	55
P1.4S05	how	27,7	0	0	0	40	15
P1.4S05	this	16,6	0	0	0	10	10
P1.4S05	argument	33,3	0	40	10	20	15
P1.4S05	develops	66,6	40	50	10	40	60
P1.4S05	on	22,2	0	0	0	10	5
P1.4S05	a	16,6	0	0	0	10	5
P1.4S05	single	61,1	40	30	0	50	20
P1.4S05	currency	100	60	60	10	70	60
P1.8S05	what	0	20	10	10	0	10
P1.8S05	we	0	0	0	0	0	5
P1.8S05	want	5,5	0	10	0	0	0
P1.8S05	to	0	0	0	0	0	0
P1.8S05	do	0	10	0	0	0	0
P1.8S05	is	0	0	0	0	0	5
P1.8S05	actually	5,5	20	0	20	10	5
P1.8S05	get	33,3	10	0	20	0	0
P1.8S05	in	88,8	60	60	30	100	5
P1.8S05	there	33,3	10	10	30	30	35
P1.8S05	and	0	0	0	0	0	5

P1.8S05	represent	88,8	70	80	40	80	40
P1.8S05	our	5,5	0	0	0	0	0
P1.8S05	country's	0	0	0	20	0	10
P1.8S05	interests	16,6	0	10	10	10	25
P1.9S04	now	11,1	0	0	0	0	5
P1.9S04	it's	0	0	0	0	0	20
P1.9S04	slow	94,4	60	100	20	100	55
P1.9S04	progress	16,6	20	10	10	10	20
P1.9S04	but	5,5	0	0	0	0	0
P1.9S04	they	5,5	0	0	10	0	10
P1.9S04	are	100	70	40	60	100	30
P1.9S04	making	16,6	0	10	10	0	5
P1.9S04	progress	27,7	30	10	10	10	25
P2.1S02	Did	6,25	0	0	0	0	5
P2.1S02	you	6,25	10	40	0	0	0
P2.1S02	do	12,5	10	0	0	0	5
P2.1S02	that	12,5	10	20	10	0	5
P2.1S02	last	87,5	90	70	50	60	50
P2.1S02	year	12,5	0	10	10	10	40
P2.1S02	as	6,25	0	10	0	10	10
P2.1S02	well	37,5	20	40	0	10	55
P2.4S01	and	0	0	0	0	0	0
P2.4S01	it	0	0	0	0	0	5
P2.4S01	was	0	0	0	0	0	5
P2.4S01	absolutely	75	90	60	10	60	40
P2.4S01	freezing	93,5	80	40	90	50	40
P2.4S03	and	0	0	0	0	0	10
P2.4S03	the	0	0	0	0	0	0
P2.4S03	pilot	12,5	0	20	10	20	10
P2.4S03	said	0	0	0	10	0	0
P2.4S03	well	6,25	10	30	0	0	10
P2.4S03	there's	0	0	0	0	0	10
P2.4S03	a	0	0	0	0	0	10
P2.4S03	ten-degree	93,7	70	50	10	20	5
P2.4S03	difference	31,2	30	10	0	0	10
P2.4S03	between	0	0	0	10	0	15
P2.4S03	Marseille	25	20	10	20	10	30
P2.4S03	and	0	10	10	0	0	25
P2.4S03	and	0	0	0	0	0	35
P2.4S03	and	0	0	0	0	0	20
P2.4S03	London	18,75	20	40	30	10	20
P2.7S04	I	6,25	0	0	0	0	0
P2.7S04	think	12,5	0	10	0	10	15
P2.7S04	people	18,7	0	10	0	10	25
P2.7S04	who	6,25	0	0	0	10	0
P2.7S04	learn	12,5	0	0	10	10	15

P2.7S04	a	6,2	0	0	0	0	5
P2.7S04	language	12,5	0	0	0	20	15
P2.7S04	just	93,7	70	50	20	40	10
P2.7S04	by	0	10	0	0	0	0
P2.7S04	picking	37,5	30	10	0	10	10
P2.7S04	it	31,2	0	0	0	0	5
P2.7S04	up	50	20	10	0	0	15
P2.7S04	you	0	30	10	0	10	0
P2.7S04	know	0	30	20	0	20	5
P2.7S04	as	6,25	30	10	0	0	5
P2.7S04	adults	68,7	60	30	10	50	30
P2.7S04	then	0	0	0	0	0	5
P2.7S04	they	0	0	0	0	0	0
P2.7S04	probably	0	0	10	0	0	5
P2.7S04	do	68,7	30	10	0	50	0
P2.7S04	pick	0	0	10	0	0	0
P2.7S04	up	0	10	10	0	0	0
P2.7S04	all	0	10	30	0	0	0
P2.7S04	sorts	12,5	10	10	0	0	0
P2.7S04	of	0	0	10	0	0	0
P2.7S04	mistakes	12,5	20	20	0	10	20
P3.1S04	It's	0	0	0	0	0	5
P3.1S04	not	18,7	10	0	50	10	45
P3.1S04	even	93,7	80	90	10	30	20
P3.1S04	been	18,7	0	0	0	0	15
P3.1S04	discussed	62,5	20	40	10	20	15
P3.1S04	till	0	0	0	0	0	0
P3.1S04	this	0	0	0	0	0	0
P3.1S04	report	12,5	10	0	0	0	10
P3.1S04	came	6,25	20	0	10	30	30
P3.1S04	out	12,5	10	0	0	0	10
P3.1S04	properly	18,7	30	40	0	10	30
P3.2S03	or	0	0	0	0	0	10
P3.2S03	would	0	0	0	0	10	5
P3.2S03	you	0	0	0	0	0	5
P3.2S03	say	0	0	0	0	0	0
P3.2S03	domestic	62,5	40	80	0	0	10
P3.2S03	violence	100	100	60	20	100	95
P3.2S03	is	0	0	0	0	0	0
P3.2S03	a	0	0	0	0	0	0
P3.2S03	private	43,7	50	40	20	0	10
P3.2S03	issue	25	10	0	0	0	40

Annexe 5e - Test perceptif sur signaux manipulés : phrases-sources

P1.1S01 (🗣️ exs.54) : DD This is a = kerfuffle today over whether the Chancellor had threatened to resign.

P1.3S01 (🗣️ exs.55) : JT : looking at it I think we're certainly going to have a better chance of snow = on Christmas Day

P1.4S04 (🗣️ exs.56) : PM : there's a perfectly sensible policy which the Conservative = cabinet the Conservative government and the Conservative Party in the House of Commons

P1.4S05 (🗣️ exs.57) : JT : and that is that we're going to wait to see = how this argument develops = on a single currency

P1.8S05 (🗣️ exs.58) : MM : what we want to do is actually get in there and represent = our country's interests

P1.9S04 (🗣️ exs.59) : PM : Now it's slow progress but they are making progress.

P2.1S02 (🗣️ exs.60) : JC : Did you do that =last year as well ?

P2.4S01 (🗣️ exs.61) : CB : and it was absolutely = freezing

P2.4S03 (🗣️ exs.62) : JC : and the pilot said well there's a ten-degree difference between Marseille and and and London

P2.7S04 (🗣️ exs.63) : VB : I think people who learn a language = just by picking it up you know as adults = then they probably do pick up all sorts of mistakes

P3.1S04 (🗣️ exs.64) : YB : It's not = even been discussed till this report came out properly









P3.2S03 (🗣️ exs.65) : YB : or would you say domestic violence = is a private issue

Contenu du CD ROM






A LIRE

BBEEdit Lite 4.6








CORPUS

-  Son
 -  BBC .wav
 -  Conversation .wav
 -  Question Time .wav
-  Transc orth
 -  BBC
 -  Conversation
 -  Question Time

CRUISE

-  fichiers .dat
-  fichiers .out
-  fichiers .ps
-  fichiers .tex
-  fichiers .tif



Etiquetage

-  Mtx
-  Mtx+phon
-  Son ()
-  TextGrid
-  TextGrid 2
-  TextGrid phonèmes











Exemples sonores ()

IPA/SAMPA














Scripts

-  PERL
 - AddPron2Matrix.pl
 - mapPhones
 - mtostat.pl
 - TextGrid2Matrix.pl
-  PRAAT
 - measure F0 for labels
 - measure intensity for labels








Sondages

-  Auditeurs
-  Instructions
-  Résultats
 -  sondage 1
 -  sondage 2
 -  sondage 3
-  Son
 -  Sondage1.wav
 -  Sondage2.wav
 -  Sondage3.wav

Synthèse voc

-  Fichiers .pho
 -  Durée
 -  F0
 -  jastradanz
 -  normal
 -  Pauses
-  Fichiers son ()
 -  Durée
 -  F0
 -  jastradanz
 -  normal
 -  Pauses

Test perceptif

-  Auditeurs
-  Instructions
-  Résultats
-  Son
 -  groupe1.wav
 -  groupe2.wav
 -  jastra.wav

Thèse.pdf